

How not to run a scientifically successful country

Early next month, the population of Switzerland will start to vote on the abandonment of research using genetically modified animals and plants. Scientists have behaved commendably but to insufficient effect.

It is not often that a country's population deliberately commits a thriving research and industrial activity to the grave. A science base can weaken through lack of investment, as is happening in the United Kingdom. It can lose competitiveness through institutional sclerosis, as has happened in France. It can collapse almost overnight as a result of political upheaval, as has happened in what was the Soviet Union. But for a country voluntarily to remove itself from a lively scientific arena in which it is highly successful is a unique phenomenon. To do so in a fully educated and democratic way would be some sort of signal sent to the rest of the world. To do so as a result of ignorance or political accident would be a tragic folly. What has gone so wrong in Switzerland that scientists and companies, faced with such a prospect in June, are taking to the streets this week and — the luckier ones at least — making contingency plans to leave the country?

Switzerland's constitution allows its people to amend it by referendum, which they do several times a year with votes decided by simple majority, and whose results are binding. On 7 June the country will vote yes or no to a ban on the production and distribution of transgenic animals, field trials with genetically modified organisms (GMOs) of any sort, and the patenting of genetically modified animals and plants (see *Nature* 389, 103; 1997). If the country votes yes, new and possibly existing activities become illegal from 8 June. The start of the referendum is even more imminent, because postal voting is allowed from 10 May.

Unsurprisingly, the government is against the ban. It recognizes the huge contribution transgenic animals such as knock-out mice are making in biology, the vigour of the Swiss biology community, and the strength and economic importance of its biology-based companies. Without access to transgenic animals, key parts of the Swiss academic and industrial research and technology base will up and leave. Senior vacancies in Swiss universities already remain unfilled pending the outcome of the referendum. Almost 500 graduate students working with transgenic animals, and 80 planning field trials with GMOs, might lose their chances of a doctorate. Public funding of overseas collaborations with transgenic animals would also become illegal.

Optimism

Several months ago, the outcome of the debate appeared evenly balanced, but researchers were optimistic. Swiss industry would weigh in and help to save the day, it was said. But all the signs are that early industrial initiatives failed. Commendably, the government offered legislation, intended to defuse the referendum, that would establish a national ethical committee, increase the period of legal liability after genetic engineering to 30 years, and tighten research regulations. Only after a belated perception that industry and government had failed to make a difference, Swiss Nobel prizewinners held a press conference, describing the proposed ban as "short-sighted and senseless", while a hundred or more laboratories have held open days.

In the meantime, the opponents of such research have been active

on quite a different level, using large posters and slogans ("We want healthy bread", "genes = danger"), and, in one highly publicized event, halting vessels carrying transgenic maize by river into the country. The result of all of this turns out to give grounds for pessimism. Too often comments have been heard from the public that exhibit a purely emotional reaction to the referendum with little consideration of its wider implications. "I'm for the amendment because I don't want human cloning" is one depressingly symptomatic comment. The signs from private opinion polls are that researchers have made little headway.

Concerns

It was always too simplistic (and is now futile) to argue, as some scientists have, that education is the key. The public has emotive responses that cannot be ignored. Yes, there is a gut reaction to some developments that can be allayed by familiarity. But there are other components that tap more deeply rooted concerns, engendered by culture and a sense of values associated with what animal or human existence is about. To attempt to distinguish between these components appears irrelevant, given the Green movement's success at enhancing every fear it can lay its hands on. For scientists to find themselves in a national battle against such fears is wrong in principle and a poor reflection of the Swiss way of conducting the country's affairs. Be that as it may, scientists estimate that as much as 40 per cent of the population has little idea of what the real issues are and is beyond the reach of anything scientists can do. The fact that major pharmaceutical companies are only grudgingly respected by many Swiss does not help.

In these circumstances, what can be done? The scientists have no choice but to conclude that attempts to enlighten the public have proved inadequate. Researchers are right to lower the level of debate and take to the streets, as they have done this week, waving placards, though perhaps their level ("Don't repeat the example of Galileo") is still sometimes overly cerebral. In the short time that is left, researchers must try even harder to communicate the message that their work must be allowed to continue, above all, because of its potential contribution to human health. Appeals to scientific freedom, or to the potential threat to scientists' careers, are unlikely to make much impact.

The other message to emphasize is that, where there are fears, whether or not these are considered justified by the scientific community, they can be addressed through a system of regulation and monitoring — as the government has already proposed, and other countries, such as Germany, have demonstrated.

But, above all, the current stalemate should lead the Swiss people to reconsider the appropriateness of decision-making by referendum in a highly complex world where few can have detailed knowledge of the issues at stake and their full implications. If Switzerland does vote in June to cut itself off from one of the most potentially valuable channels of modern biomedical research, the blame should be apportioned more widely than to a group of highly organized activists. □



Montserrat residents 'lost faith' in volcanologists' warnings

[LONDON] Scientists on the Caribbean island of Montserrat lost credibility in the eyes of the island's inhabitants by wrongly predicting when its volcano would erupt during 1996 and 1997, according to a survey of the island's residents.

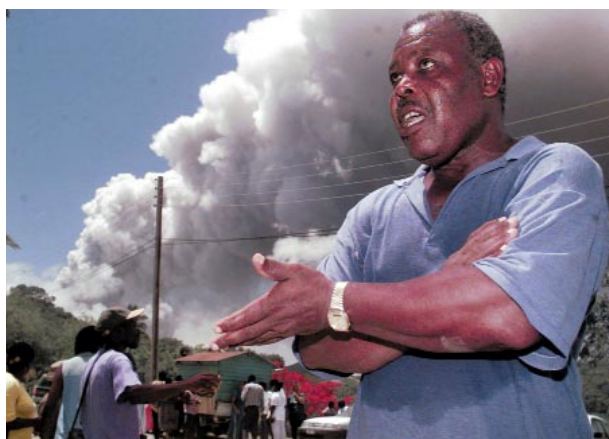
The survey is due to be published within weeks, and could have a significant impact on debates about the role of science communication in disaster management policy.

David Sanderson, a researcher at the Oxford Centre for Disaster Studies and co-author of the survey, says the island's residents were disappointed when scientists made mistakes, and were confused by their use of complicated jargon when explaining the volcano's activity.

Sanderson says that the scientists were not only expected to know in advance when the volcano would erupt, but also to communicate this information in layman's terms.

The British government commissioned the survey in the aftermath of major volcanic activity last year. More than 20 people died trapped in their homes, despite daily warnings to leave the area. Preliminary findings were revealed at last week's annual meeting of the Geological Society of London.

Stephen Sparks, professor of volcanology at the University of Bristol and chief scientist at the Montserrat Volcano Observatory, accepts that mistakes were made, and that scientists working in disaster situations need help communicating risk information.



Heated debate: a resident of Montserrat discusses a deal offered to inhabitants of the island by the British government last summer, as ash and steam billow from the Soufriere Hills volcano behind. Volcanologists on the island, criticized by residents for their poor forecasting of eruptions, say it proved difficult to convince those who have lived safely in an area all their lives that it had become dangerous.

"Volcanology is an uncertain science," he told the meeting. "There are aspects we do not understand. But we are not trained to communicate doubt and uncertainty." Sparks added that scientists need feedback from the public as well as "help from people like sociologists and disaster managers".

The findings of the survey also have implications for the role of scientific advice to governments involved in disaster situations. Some scientists at last week's meeting said privately that their task of communicating information about risks was further complicated by the lack of a disaster preparation plan, as well as by politicians' unrealistic demands on their expertise.

Montserrat is a small island 11 miles long. The volcano is in the south, and has been

quiet for much of this century. Volcanic activity was renewed in July 1995. Much of the south, where most people lived, is now covered in ash following the 1997 eruption. Most of the island's 12,000 residents have been evacuated; the rest are now in the safer north. The volcano is now relatively quiet.

Local criticism of the observatory stems from the scientists' patchy record in predicting eruptions, and the direction of the flows of hot ash and magma — known as pyroclastic flows. The observatory's staff failed to predict the explosion of magma on 17 September 1996, which caused a 40,000-foot high plume of ash, and deposited 600,000 tonnes of ash on the south of the island.

Their prediction of a similar eruption in December the same year turned out to be a false alarm. Residents who had been evacuated were sent back to their homes.

Six months later, more than 20 people died in and around their homes during a third major eruption in June 1997. Sparks says an evacuation order warning of this eruption had been given months in advance. "We don't really understand why people were still in there against official advice." One possible reason, says Sanderson, is that people were reluctant to move to the less developed north of the island.

Sanderson says that many of the island's residents said they felt the scientists' prediction record was sometimes no better than their own. Islanders said that generations of residents had monitored the volcano all their lives, and had observed changes in rockfalls, pyroclastic flows, sea level, humidity and animal behaviour.

They wanted scientists to tell them more than they already knew. "I still believe that the scientists do not have the capability to alert us when it's needed," said one resident. "You have to alert yourself."

The survey also revealed that there was

Australian research centres escape axe

[SYDNEY] After a long battle, the future of Australia's Cooperative Research Centres (CRCs) seems to have been secured. The latest review of the scheme has led to the Coalition government committing annual funding at the reduced level of last year — A\$138 million (US\$104 million) — and approving a round of applications for new centres, in competition with the 35 existing ones which are eligible for renewal.

The CRC scheme is widely seen as a successful strategy linking government, university and industry researchers on tightly focused tasks, mostly with medium-term, commercial goals. Successful centres have previously had to raise substantial funding from partner organizations to supplement a government grant of about A\$2 million a year for seven years, with the possibility of a second seven-year term.

But the scheme, established by the previous Labor government, had recently

appeared vulnerable under the conservative Coalition, with cuts in last year's budget and a drastic call from an industry review for a 70 per cent cut in government support (see *Nature* 387, 222 & 388, 507; 1997).

The decision, announced last week by the science minister, John Moore, was greeted with relief by the heads of the 67 CRCs, who had mounted a sustained lobby to defend the scheme. But the government has not yet released details of changes recommended in the latest review, 'guided' by the chief scientist, John Stocker, and former banking chief Don Mercer. This is now a focus of some lingering concerns.

Paul Wellings, the official in charge of science and technology in Moore's department, says there will be more emphasis on the commercialization of research. "The CRCs [need] very good business plans and a clear sense of what they are going to deliver," he says. **Peter Pockley**

widespread confusion about the eruption warning system put in place by scientists and the local authorities, as well as tensions between scientists and the authorities over the nature of this warning system.

The system used sirens to alert residents of an impending eruption, daily newspaper and radio reports by scientists, and maps agreed by government and scientists indicating the risks in different areas.

But survey respondents voiced dissatisfaction with most of these methods, and relied instead on information passed by word of mouth. The radio reports were considered by some to contain too many technical words. And the 'risk maps' introduced after the September 1996 eruption to educate people about the level of risk caused much confusion, and had to be simplified.

The early maps divided the island into seven zones, each with one of six different levels of risk, or 'alert level'. The maps were updated six times as data changed.

Sanderson says this level of detail baffled residents, and the maps were simplified to just three zones as the volcano became more active. Many islanders were unaware of the existence of the earlier maps. Others did not know which zone they lived in. Some were unaware of changes to the status of their zones. "Eighty-eight per cent of respondents in a zone previously considered 'unsafe' did not know that this had changed to a zone considered highly dangerous," says Sanderson.

"All this talk of zoning is confusing the man on the street," said one respondent. "All these words such as 'progressive gravitational induced collapses', and pumice falls and all this — what does it mean?"

The risk maps had to be cleared by the authorities, and scientists and politicians had differing views on what they should contain. Richard Robertson, of the Seismic Research Unit at the University of the West Indies, says that scientists favoured simpler maps. But the authorities insisted on 'micro-zonation', as they did not want to evacuate the whole of the south, even though the scientists themselves were not sure of the accuracy of their predictions.

The question of whether scientists could forecast the timing and direction of pyroclastic flows to within a margin of error of metres is a "moot point", says Geoff Wadge of the University of Reading, who also worked at Montserrat. He says scientists were never forced to do anything. But he acknowledges "pressure to do micro-zonation".

Wadge says that, despite its failings, micro-zonation at Montserrat had its uses. He says that the authorities there have a difficult job. They need to balance the desire not to destroy a stable community with that of public safety. A complete evacuation and no eruption would have damaged the economy, but allowing people to live as normal might have led to many more deaths. **Ehsan Masood**



Hot wired: the new Abilene network will form a high-speed data communication backbone serving the main research universities in the United States. Fibre-optic cables will run mainly under railway lines, where they will be relatively safe from accidental disruption.

Faster Internet system will overcome congestion

[WASHINGTON] Research universities in the United States will gain access to a new and immensely powerful Internet upgrade early next year. It should enable them to by-pass congestion on the existing Internet and conduct experiments that involve a far larger volume of data transfer than is now possible.

Under an agreement announced at the White House in Washington last week by Vice-President Al Gore, the new network will be built and operated by private contractors. It will serve as a backbone network for Internet2, a project involving 160 US research universities that have combined to arrange faster links than are at present available from commercial suppliers.

The network, called Abilene, is being built by Qwest Communications — a telecommunications corporation based in Denver, Colorado — together with two technical partners, Nortel and Cisco. Most of the backbone will run under railway lines, where its fibre-optic cables are relatively safe from accidental disruption.

When it comes on line at the end of this year, the network will allow data transfer at up to 2.4 gigabits per second, later rising to 9.6 gigabits per second. The existing very high performance Backbone Network Service (vBNS), which the telecommunications company MCI operates between 92 research universities for the National Science Foundation (NSF), runs at 600 megabits per second. Most domestic and commercial users of today's Internet are lucky to get 100,000 bits per second in the United States, say experts.

Qwest and the other commercial suppliers say they do not expect Abilene to generate substantial direct revenues from the research universities using it. According to Joseph Nacchio, president of Qwest, the university network will be useful to his corporation mainly as a test-bed for future commercial networks, and also as an entrée to telecommunications business from the universities.

Analysts expect these new research net-

works to serve as precursors for the public Internet of the future, paralleling the way NSFnet, the network that NSF operated for the universities in the early 1990s, foreshadowed the existing public Internet.

Gore extolled the new agreement as "a startling advance" and predicted that advances in voice interaction technology would soon make the World Wide Web available to "hundreds of millions of people". Researchers will use it for real-time access to supercomputers and to construct Internet applications in fields such as 'telemedicine' and industrial control that need prompt, reliable delivery of information.

The Internet2 consortium was established last year by research universities worried about severe congestion on the Internet. Each of the 160 institutions has paid \$25,000 to join the consortium and pledged to spend at least \$500,000 upgrading its own systems to link with the backbone networks.

Abilene will run alongside vBNS, but will run faster and is due to continue operating after NSF's support for vBNS expires in the year 2000. US government agencies are already researching even faster links under the Next Generation Internet initiative. Gore also announced last week the award of 27 major research grants by the Defense Advanced Research Projects Agency, worth \$50 million over three years, to develop applications for this still-faster network.

Some observers believe the availability of these high-speed links will lead to new pricing structures that will directly charge their users. "We are going to see a replay of the situation we used to have with mainframe computers, where you pay for what you use," predicts Stephen Wolff, a former network manager at NSF who now works for Cisco.

The deal suggests that universities may get cheap access to fast networks from suppliers interested in the much larger market that will emerge when industry and business start using the technology. **Colin Macilwain**

Call for Europe-wide nuclear physics push

[MUNICH] Nuclear physicists in Europe must collaborate more closely to answer the next big questions in their fields, according to experts from 23 research organizations.

A new report published by the Nuclear Physics European Collaboration Committee (NuPECC), an advisory body to the European Science Foundation, says it is important to draw together Europe's somewhat disparate nuclear physics community because the subject has reached a point at which large and expensive equipment is required to take it further.

One of the report's main recommendations is that the options for a second-generation radioactive ion beam in Europe should be defined as soon as possible. Such a beam would be able to create heavier ions, rich in neutrons, than has so far been possible. It would also allow the structure and characteristics of heavy ions to be determined, thus providing a better understanding of the first milliseconds of the Universe during which thousands of different types of such ions are predicted to have existed.

Sydney Galès, chairman of NuPECC and director of the Institute of Nuclear Physics in Orsay, says that it is vital for Europe to move

quickly in heavy-ion physics because, although it currently plays a leading role in this field, it could lose its competitive edge as the United States and Japan both implement plans to upgrade their equipment.

The technologies involved in the design of future radioactive heavy-ion beams, such as high-density accelerators or reactors, share characteristics with technologies being developed for other uses, such as transmutation techniques for disposing of nuclear waste, and next-generation neutron spallation sources.

As a result, NuPECC has set up an interdisciplinary study group to identify options. The group will report at the end of next year, and at that stage NuPECC will put out a call for funding to research agencies, says Galès.

The report also recommends that a high-luminosity 25-GeV electron beam facility be built in Europe soon, to further studies on the structure and dynamics of hadrons. All the major hadron experiments now being conducted in Europe will be completed in the next six or so years.

The scientific council of the German particle-physics research centre DESY agreed last autumn to consider using its planned high-energy TESLA linear collider to

generate such an electron beam, which could inject electrons at the appropriate energy into its existing HERA electron storage ring.

This strategy would be considerably cheaper than building a 25-GeV electron facility from scratch. DESY has asked NuPECC to provide a clear physics case for the idea, and to define a clear user community and a costing within the next two years.

NuPECC's report is the second of a planned series of six-yearly surveys. According to Galès, three-quarters of the recommendations of the last report, published in 1991, were put into practice. He hopes that the current report will be equally well received by funding agencies, although admitting that "the political climate has changed in the last few years".

Some of the report's recommendations are already being implemented. For example, the European particle-physics laboratory CERN has approved a heavy-ion experiment called Alice (A Large Ion Collider Experiment) to run on its Large Hadron Collider, which comes into operation in 2005. This will help to answer questions about properties of quarks and gluons, the building blocks of nuclear particles.

Alison Abbott

French research agency to seek ruling on Holocaust sceptics

[PARIS] France's national scientific research agency is to try to clarify a course of action on researchers who actively promote 'revisionist' arguments such as denying the existence of the Nazi gas chambers.

The Centre Nationale de la Recherche Scientifique (CNRS), which is split between the need to preserve academic freedom and a desire to discipline such individuals, will ask its ethics committee, COMETS, for advice, on which the agency's national committee will then act.

The question of revisionism in French academic circles has been kept simmering for some time by the revisionist activities of Serge Thion, a CNRS researcher, as well as those of several other scientists.

The issues are particularly sensitive given that France has recently been reminded of Fascist activities in its past by the trial of Maurice Papon — a senior civil servant imprisoned for collaboration in Nazi deportations — and the significant role of the National Front in recent regional elections.

Matters were brought to the boil in February by allegations in the German newspaper *Berliner Zeitung*, attributing revisionist arguments — in particular that the existence of the Nazi gas chambers had not been proved — to Gabor Rittersporn, a CNRS researcher working at the prestigious Marc-Bloch Franco-German



centre for social sciences in Berlin.

Earlier this month, the researcher won a court case against the newspaper and cleared his name. Rittersporn, who was born in Hungary, had denied the allegations and condemned revisionism, telling the court that some members of his family, itself part-Jewish, had perished in the Holocaust.

The allegations against Rittersporn followed revelations about his membership in the 1970s and 1980s of extreme left-wing groups that favoured free expression for revisionists. He was also a member of the editorial board of *La Vieille Taupe*, a French publisher that has published revisionist texts, including one on the gas

chambers which Thion edited.

But Rittersporn has argued that he was at the time naively defending freedom of expression, and ceased participating in these activities when he learnt they were being used to support revisionism. Etienne François, the director of the Marc-Bloch institute, said at the time that Rittersporn's past activities had been fully discussed at CNRS before his appointment, and that the agency had been satisfied by Rittersporn's explanations.

In response to the highly publicized controversy generated by the Rittersporn case, CNRS set up an expert panel chaired by Edouard Brezin, president of its executive board, to consider the issue of revisionism among academics. But Brezin says he quickly realized that such a panel had few powers to investigate the behaviour of individual researchers.

Given this lack of power, and that the urgency has dissipated with the intervening court ruling clearing Rittersporn, CNRS has decided to dissolve the panel and pass the issue to COMETS for guidance. It will then ask the CNRS's national committee, which has the power to evaluate researchers, to take appropriate action.

Brezin says he hopes COMETS will clarify the limits of academic and individual freedom for public researchers, as well as CNRS's disciplinary powers.

Declan Butler

Auto project has veered off track, says academy panel

[WASHINGTON] A joint initiative between the US government and the car industry to develop more economical cars is putting too much effort into 'hybrid' vehicles that will be costly to build and maintain. That is the conclusion of a panel of the National Research Council (NRC), part of the National Academy of Sciences complex.

The Partnership for a New Generation of Vehicles (PNGV) programme decided earlier this year to pursue hybrid vehicles as the most promising route to developing a prototype passenger car that will do 80 miles per gallon (m.p.g.) by the year 2004.

PNGV is a flagship of the Clinton administration's technology policy, and involves government laboratories working in collaboration with researchers at Ford, General Motors and Chrysler. It chose hybrid electric vehicles, which can switch between a diesel engine and a battery depending on conditions, because more sophisticated technologies, such as fuel cells, will not be ready by 2004.

But the hybrid approach may never be economical, says the NRC panel in its fourth annual assessment of the project. This is because of the extra costs of building and maintaining a car with two power systems. Despite the better fuel economy of a hybrid vehicle, the panel says, "the difference is probably not enough to offset the higher cost" of building and maintaining it.

The panel was chaired by Trevor Jones of Echlin, a car component supplier based in Cleveland, Ohio. It says that PNGV partners should put more effort into technologies with greater long-term potential for big improvements in fuel economy, such as fuel cells. It adds that, while awaiting such breakthroughs, the partners should try to develop a non-hybrid vehicle with a more realizable fuel economy target of 60 m.p.g.

The panel suggests that some of the project's effort should be directed to increasing the fuel economy of light trucks and sport-utility vehicles, which together now account for half of the car market in the United States. The project is currently focused on increasing the economy of a standard four-door family car, such as the Ford Taurus.

William Daley, the commerce secretary, and Federico Peña, the energy secretary, issued statements that welcomed the NRC report but failed to address the issues it raises. But Eric Clark, a spokesman for the Department of Commerce, said that PNGV was only four years into a ten-year project, and that it would now start to focus more on questions of cost. "The affordability issue is now at the forefront of our work," he said.

C.M.

US universities under fire for 'antiquated' teaching

[WASHINGTON] Research universities in the United States consistently ignore the needs of their undergraduate students, failing to let them do research or even to find out what it involves, says a report released this week.

A commission set up by the Carnegie Foundation for the Advancement of Teaching to investigate undergraduate education also finds that the leading universities provide undergraduates with few of the communication or other skills needed in a job.

The commission finds that the research universities — which award about half of all undergraduate degrees in science and engineering in the United States — continue to teach science using methods borrowed from liberal art schools a century ago and clumsily adapted to the massed ranks of undergraduates now accepted each year.

It finds that the undergraduates, as well as being badly taught, are largely excluded from the intellectual life of the universities. They are, in the commission's words, "second-class citizens who are allowed to pay taxes but barred from voting, the guests at the banquet who pay their share of the tab but are given leftovers". It blames incentive schemes that encourage faculty members to value research above teaching, as well as general resistance to change in university departments.

The commission started its work under the tutelage of Ernest Boyer, former president of the Carnegie Foundation, and continued after his death under the leadership of Shirley Kenny, president of the State University of New York at Stony Brook.

Its criticisms of the universities, which resonate with growing resentment in the United States over the fees they charge

undergraduates, received intense media coverage when they were released.

But the commission's recommendations for change promise little relief for the bank balances of students. Top of the list of ten changes that it says the universities should make is the replacement of lecturing of massed ranks of first-year students with



Alberts: changes are already under way.

enquiry-based methods, including personal attention through seminars. "The focal point of the first year should be a small seminar taught by experienced faculty," the commission says.

This personal attention should continue through mentoring throughout the degree course, it suggests. Departmental boundaries should be breached to encourage interdisciplinary work, which "must reflect students' needs rather than departmental interests". Communication skills should be integrated into teaching, the commission says, noting that "the failure of research universities seems most serious in conferring degrees upon inarticulate students".

Bruce Alberts, president of the National Academy of Sciences and a member of the commission, says progress is under way to address many of its findings. At Stanford University in California, he says, first-year students take at least one enquiry-based course, and staff promotion is considered by a committee with a member who assesses teaching performance.

Colin Macilwain

UK biotech flagship meets stormy waters

[LONDON] British Biotech, one of the leading UK biotechnology companies, last week sacked its director of clinical research for allegedly passing confidential information to two of the company's investors.

Andrew Millar was dismissed for allegedly disclosing doubts about the company's commercial strategy and the status of trials of its flagship drugs, zacutax for pancreatitis, and marimastat for cancer.

But the sacking did little to reassure analysts that the troubled company was about to turn the corner. The company's stock market value has plummeted from £1.9 billion (US\$3.2 billion) last year to £372 million last week following a series of internal controversies, including a delay in reporting interim results of drugs trials.

British Biotech is also being investigated by the US Securities and Exchange Commission over the accuracy of press statements about marimastat. Optimistic assessments of this drug led to large increases in the company's share price.

It has also emerged that two of the company's senior officials sold shares for a profit of around £1.2 million before the announcement of problems with batimastat, its previous lead cancer drug that was abandoned in 1995.

A senior fund manager at Perpetual, which holds 8 per cent of British Biotech shares, is reported to have said that Millar was concerned that "the status of the trials was completely at odds with the [company's] grand strategy".

Ehsan Masood

Row erupts over child aggression study

[WASHINGTON] Psychiatrists under attack for using the drug fenfluramine in a trial examining aggression in poor, non-white, inner-city boys last week defended themselves, along with those who gave ethical approval to the experiments, against criticism appearing in the press.

Critics of researchers at the New York State Psychiatric Institute (NYSPI) in Manhattan claim that a paper published last year in the *Archives of General Psychiatry* shows the research was in flagrant breach of federal ethics rules governing research on children.

But the researchers say they obeyed those rules, and that the critics are using specious objections to cast doubt on the proper scientific investigation of aggressive behaviour. "The implication is that you should not study the relationship between aggression and biology," says Daniel Pine, a child psychiatrist and lead author of the paper, who describes the research as critical for helping to solve "a major public health problem".

B. Timothy Walsh, chair of the institute's Institutional Review Board (IRB), which approved the study — and is now under fire for having done so — says the panel correctly sought to interpret government ethics rules in the light of the facts of the protocol, not how it might be seen by critics.

"We feel we're being singled out somewhat unfairly," says Walsh, pointing out that other IRBs have approved studies using fenfluramine in children, and that the federal government regularly funds such research.

Pine and his co-authors reported a positive correlation between an indicator of fenfluramine-induced activity of serotonin in the central nervous system and both aggressive behaviour and social circumstances conducive to the development of such behaviour. The 34 boys studied, aged six to ten and either black or Latino, were the younger brothers of juvenile delinquents traced by investigators through court records.

Fenfluramine — part of the controversial diet drug combination 'Fen-Phen', which was withdrawn last year after it was discovered to cause heart valve damage — increases serotonin levels in the central nervous system. The New York trial, funded largely by the Lowenstein Foundation, with some support from the National Institute of Mental Health, was halted in 1995, two years before the drug was withdrawn.

Last week, the Long Island newspaper *Newsday*, followed by *The New York Times* and the *New York Post*, ran articles on the study quoting criticism from ethicists and others. The Office for Protection from Research Risks at the National Institutes of Health (NIH) has since confirmed that it is investigating protection of human subjects at the psychiatric institute.

For non-therapeutic research involving a "minor increase over minimal risk" — the category in which the IRB at the psychiatric institute considered the trial — federal rules say that children may be subjected only to experiences that are "reasonably commensurate" with their actual or expected medical situations. Also, the study must be likely to produce 'generalizable' knowledge about the subject's "disorder or condition" which is essential for understanding or improving it.

Disability Advocates, a public-interest law office in Albany, New York, which complained about the study to the NIH, contends that the children underwent experiences — including overnight fasting, intravenous catheterization and oral administration of fenfluramine — beyond those that medically healthy children might reasonably expect.

It adds that because the children were healthy — a prerequisite for participation — the study could not have yielded vital, generalizable knowledge on their "disorder or condition". Cliff Zucker, executive director of Disability Advocates, argues that researchers, the federal government and IRBs have all ignored "the clear rules about what's permissible and what's not".

But both Pine and Walsh say that many of the children were psychiatrically disturbed, and that all were at "high risk" of developing antisocial behaviour or aggression. This, they say, constitutes a "condition" under the federal rules.

Walsh argues that, in the IRB's view, "going to a medical setting, getting stuck

with a needle and taking a pill that may make you feel a little funny" are "commensurate" with the experience of children without serious medical illnesses. Pine and Walsh say the IRB took care in approving the trial, referring it to a subcommittee with expertise on children and requiring extra documentation about fenfluramine's risks.

But others say the study had elements that should have worried the IRB. One was the racial composition of the group, which was 44 per cent black and 56 per cent Hispanic, a "highly disturbing" skew, according to Arthur Caplan, director of the Centre for Bioethics at the University of Pennsylvania.

Racial balance is particularly important in research with no direct benefit to subjects and that is of more than minimal risk, says Robert Levine, a professor of medicine who chairs Yale University's IRB.

The investigators have argued that their neighbourhood of New York City is largely Hispanic and black, and that they were seeking funding to expand recruitment to include whites when the trial was stopped for unrelated reasons.

Pine admits that the lack of white participants "was a flaw in our study", but only because it limited the generalizability of the findings. Caplan argues that potentially contentious work of this nature "will need to be taken to a different review level than the local institution". Such trials, he says, should be subject to a review body that "has not got the same pressures on it as the local IRB to get it approved".

Meredith Wadman

Rockets launch Spain's eye on the sky



Fireworks lit up the sky last week at a ceremony to inaugurate a futuristic-looking planetarium in Valencia, eastern Spain. The planetarium, in the form of a human eye, was designed by Valencian architect Santiago Calatrava and forms part of a new art and science complex.

AP/FEI/J. C. CARDENAS

Police violence flares up at Russian education protest

[MOSCOW] Police in Ekaterinburg, the capital of the Sverdlovsk district in the Urals and home of the Russian president, Boris Yeltsin, last week attacked students from local universities who had taken part in a national protest against proposed educational reforms (see *Nature* 392, 429; 1998).

Between 3,000 and 4,000 students gathered at the city's Youth Palace for a meeting that had been approved by the city council. After discussing their problems and the apparent indifference of federal and local authorities to their needs, they adopted a resolution demanding that the president and the cabinet abandon reforms that, they argue, would make higher education unaffordable for most people.

The leaders of the students' trade-union association, which organized the meeting, then asked the students to return home. But more than 1,500 students, angry that no member of the city authorities had agreed to meet them, moved to the Ekaterinburg city council.

They then marched to the Sverdlovsk district administration, where 200 students approached the main entrance of the district government's building. Some were eventually admitted, and met Anatoly Gaida, the first deputy head of the district government, responsible for education, science and culture, who endorsed the students' demands and added some of his own.

Meanwhile, more than 250 members of the Special Police Force, armed with steel shields and rubber clubs and accompanied

by a military armoured car, started to move the students out of the government building. Injuries were received by many students.

Local television reports, subsequently shown nationwide, generated wide indignation. Yeltsin ordered the newly appointed acting minister of internal affairs, Sergei Stepashin, to investigate the case urgently, and the parliament also set up a special commission of inquiry.

Sergei Kirienko, the acting prime minister, said the cabinet was deeply concerned by the incident. General Boris Gromov, a former head of Russian troops in Afghanistan, said that "only silly and irresponsible people could throw at our future with police clubs".

Yuri Brunsitsin, the president's recently appointed representative in the Sverdlovsk district, promised to pass on the students' demands to Yeltsin. The students told him that, although they were until recently "passive politically", that did not mean their needs "should be neglected by the authorities".

University lecturers from Ekaterinburg who attended the student meeting told local officials that, although they had previously acted as a "separating layer" between the authorities and the young people, they would join the students "if the present educational policy is kept unchanged".

Three high-ranking officials of the Sverdlovsk district government, including Gaida, have already been dismissed. The heads of the district and city police forces, who ordered the use of the special forces, are also likely to lose their jobs.

Carl Levitin

Japan set to inject funds into endocrine disrupter research

[TOKYO] Research into the effects of endocrine disrupters, the man-made chemicals suspected of disrupting human reproductive functions, is about to expand in Japan. Such research is likely to receive a significant boost as part of the government's 16 trillion yen (\$121 billion) economic stimulus package for the current fiscal year (which began on 1 April).

Plans to allocate ¥10 billion towards setting up an endocrine disrupter research laboratory were announced by Japan's Liberal Democratic Party (LDP), the ruling party, last week. If approved by the parliament, this would become the largest national project on 'environment hormones', as the chemicals are known in Japan.

Details of the stimulus package are due to be announced by the Finance Ministry on 24 April. Although the issue of endocrine disrupters is not directly relevant to the government's economy-boosting measures, it is considered "an urgent matter that must be tackled as part of the national agenda", according to an LDP spokesman.

Widespread concern about the effects of 'environment hormones' was precipitated last year by the release of a report compiled by the Environment Agency. This listed 67 chemical compounds, such as dioxins, polychlorinated biphenyl compounds (PCBs) and organic tin compounds, which were suspected of mimicking naturally occurring sex hormones.

Although the details of endocrine disrupters' effects on the human reproductive system are still unknown, many scientists agree on the importance of increasing knowledge about the potential danger of man-made chemicals being dispersed into the environment.

"It is not easy to determine the effect of each chemical substance on mankind that has been released into the environment," says Taisei Nomura, a professor in embryology at Osaka University. "But, given their potential impact on human reproductive organs, it is essential that more research is carried out."

The new laboratory would be set up at the Environment Agency's National Institute for Environmental Studies, in Tsukuba. It would focus on methods to detect the presence of endocrine disrupters by analysing chemical compounds found in soil and water samples.

Meanwhile, the Science and Technology Agency intends to fund a separate three-year research programme on endocrine disrupters and other environmental pollutants, such as formaldehyde. This project, the cost of which is estimated to be more than ¥400 million a year, will be carried out jointly by 30 national institutes and industry.

Asako Saegusa

Heed public opinion, British geologists told

[LONDON] More of Britain's publicly funded Earth sciences research should be relevant to the needs of people and policy-makers, according to a former leader of the country's natural environment research body.

Sir John Knill, who used to head the Natural Environment Research Council (NERC), told the annual meeting of the Geological Society last week that informal research showed the public had a high level of interest in the applied Earth sciences, such as climate change and natural hazards.

But whereas such topics are a high priority in NERC's directed research programmes, data on 'responsive mode' funding show that Earth scientists are more interested in more specialized fields. "There is a mismatch here," he said. "We need to convince ourselves that public opinion matters."

In a keynote lecture, 'How we need to



Knill: Earth scientists should consult public.

convince society that it needs geological research', Knill told his audience that by paying less attention to public interests, Earth scientists had lowered their stock in the eyes of policy-makers. He revealed that the Geological Society is one of the

few learned bodies not being consulted in the second round of the government's Technology Foresight programme.

Knill said that if Earth scientists want to be heard in government, they need also to lobby civil servants. "What a minister writes or says is usually written by other people. We need to influence those 'other people'."

Ehsan Masood

Data glitches delay observation satellite

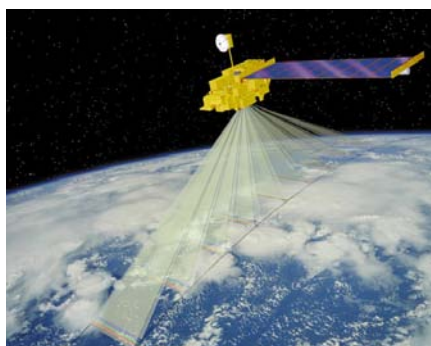
[WASHINGTON] Software problems will delay the long-awaited launch of the first major satellite in the US space agency NASA's Earth Observing System (EOS), a network of at least three large spacecraft and many smaller ones designed to study the Earth as an integrated system.

The AM-1 spacecraft — which carries US, Japanese and Canadian sensors for monitoring clouds, ice, land and water — was to have been launched in June, but will now remain grounded until at least December.

The problem is with a segment of the troubled EOS data system known as the EOSDIS, one of the most complex software engineering projects ever attempted. The \$2 billion EOSDIS is designed to handle unprecedented volumes of data from a variety of orbiting sensors over a decade or more, and to format the data for a wide range of users, from scientists to the general public.

Although the system passed an important test of its science data processing segment last summer (see *Nature* 389, 108; 1997), other difficulties have continued. One problem area is the software's Flight Operations Segment (FOS), designed to control and monitor the health of all EOS spacecraft and to schedule onboard activities such as instrument pointing.

NASA had hoped that bugs in earlier versions of the FOS would be corrected in a new version which was delivered last month by the Lockheed Martin Corporation. But some of the old problems remain, and new ones have appeared, according to EOSDIS



Still grounded: A computer-generated image of the EOS AM-1 spacecraft orbiting Earth.

project manager Rick Obenschain of the NASA Goddard Space Flight Center in Maryland.

If the FOS cannot be fixed in time for the AM-1 launch, the space agency is considering adapting commercially available software similar to that used for the recently launched Tropical Rainfall Measuring Mission and the forthcoming Landsat 7 — which also has slipped from a July launch to at least February, due to a power supply problem with its main instrument.

Any off-the-shelf alternative would not be a full replacement for the FOS, says Obenschain, and would be a one-time-only fix. It is also unlikely that it would be ready by December, he says.

The EOSDIS is “meant to be everything to everybody”, he says, which may be its biggest problem. In retrospect, says Obenschain, it was “probably not a very good intent” to design a single, massive data sys-

tem to handle all the needs of all EOS spacecraft (more than two million lines of computer code have been written already).

The project has suffered major delays, attrition rates as high as 38 per cent among its software engineers, and perhaps now an insurmountable problem with the FOS for AM-1.

Scientifically, the expected slip in the schedule will have little impact. Byron Tapley of the University of Texas, who will take over next week as chair of the science executive committee for the EOS Investigators Working Group, says there will be “no major Earth-shaking loss in science opportunity” if the delay lasts only a few months.

But the future of the EOSDIS may be on the line. Several outside advisory groups have recommended that NASA should switch to a distributed or ‘federated’ system, whereby many different organizations would supply the research community with data from one or two EOS instruments.

In December the space agency picked two dozen Earth Science Information Partners, who will have three to five years to develop and distribute prototype data products for scientists and other users.

Meanwhile the seven large, centralized data archives organized under the current EOSDIS architecture are being reviewed by a committee of the National Academy of Sciences. The committee, which is chaired by Francis Bretherton of the University of Wisconsin, meets in Washington this week, and hopes to report by the end of the summer.

Tony Reichhardt

German sex killings prompt decision to create a DNA database

[MUNICH] Germany's federal criminal office is to introduce a central DNA database for sex offenders, convicted murderers and others who commit serious crimes, including members of house-breaking rings. The minister of the interior, Manfred Kanther, approved the plan last week.

The move follows a series of largely unsolved sex killings of children in the past few months in Germany. Political pressure to introduce a DNA database grew particularly strong after the most recent case, the murder of a 13-year-old girl in Lower Saxony last month.

Rapid advances in DNA analysis and bioinformatics have made DNA databases invaluable to forensic scientists. Such databases are already used in the United States, Britain, the Netherlands and Austria, and pressure is growing for similar moves in France (see *Nature* 392, 430; 1998). But Germany has held back until now, mainly because of concern about

privacy and data protection issues.

Police in Germany already use genetic fingerprinting based on samples of hair, tissue, saliva or sperm found on the bodies or clothes of victims. But they have not had the opportunity to compare the genetic profile of such samples with data held in a central DNA database.

In the case of the recent Lower Saxony murder, police had to request saliva samples from all the 18,000 men between the ages of 18 and 30 in the region in which they believe the murderer lives to allow a comparison with the murderer's genetic fingerprint.

The samples are being analysed in police laboratories in Hanover. The procedure has been criticized for its high costs and relatively low likelihood of success, as donations are voluntary and the police are relying more on being able to rule out individuals than on identifying the murderer directly.

“Germany can no longer afford to do

without an instrument that has proved an efficient criminological tool,” says Detlef Dauke, a spokesman for the Ministry of the Interior.

The DNA database will be built up gradually. According to Dauke, the technical equipment is already available at the federal criminal office. Germany's 16 *Länder* (states), which are responsible for policing, will start to transmit existing DNA data immediately.

It remains to be decided for which crimes compulsory production of a DNA sample will be required and for how long samples will be stored. And it is still unclear whether the operation of the database will need to be backed by new legislation, if its data are to be acceptable as legal evidence in a trial.

But there is a broad consensus that the database will help to speed the rate at which serious crimes are solved, and will help to lead to the immediate ruling out of innocent suspects.

Quirin Schiermeier

US tells Russians to keep out of Iran's missile programme

[LONDON] A number of Russian research agencies have been warned by the US State Department that they risk losing millions of dollars worth of technology assistance if it is shown that their researchers have been helping Iran's missile programme.

A department spokesman last week denied a report in *USA Today* that a 'black list' had been drawn up of about 20 such agencies considered ineligible to receive support from the US government because of such involvement. But he confirmed that the government had compiled "informal lists of entities where we have concerns that they might be involved in proliferation activities".

The United States spends around \$50 million on programmes designed to steer Russia's weapons scientists towards non-military research.

Among projects the newspaper claimed have been denied US funding is one at the Baltic State Technical University in St Petersburg to apply rocket motor technology to the high-temperature destruction of chemical waste.

Max Planck institutes face closer scrutiny

[MUNICH] Germany's Max Planck Society has decided to introduce a system of six-yearly, discipline-orientated 'comparative scientific assessments' in its 80 institutes, in addition to its regular biennial evaluations of individual institutes.

Performance in a discipline will not only be compared between Max Planck institutes, but also with other institutes in the same discipline both in Germany and elsewhere. The results of the comparative review will influence the future allocation of funds, which will be shifted to institutes conducting the best research. Institutes that prove to be less efficient will get less support.

Russian science budget 'secure against cuts'

[MOSCOW] Sergei Kirienko, Russia's acting prime minister, promised in a meeting with the presidium of the Russian Academy of Sciences last week that scientific research will receive the public funding allocated to it in the approved budget "without any reductions".

Kirienko, whose nomination is currently in dispute between president Boris Yeltsin and the Duma — the lower house of the Russian parliament — also said the cabinet will pay special attention to the commercialization of scientific results. "We all need to know how much of the money

earned belongs to a scientist, how much to his institute, and how much to the state," he said. The Ministry for External Economic Relations and Commerce has been asked to look into the matter.

Strangway to lead push for innovation in Canada

[MONTREAL] David Strangway, former president and vice-chancellor of the University of British Columbia (UBC), has been appointed president of the Canada Foundation for Innovation. The foundation was set up last May by the government of Canada with initial capital of C\$800 million (US\$560 million) over five years to invest in infrastructure for research and development in not-for-profit research institutions.

Strangway is credited with developing UBC's research capabilities. He worked previously as a geophysical specialist for the United Nations, with major international companies, and with the US National Aeronautics and Space Administration, where as chief of the geophysics branch he was responsible for the geological aspects of the Apollo missions.

He became chair of the geology department at the University of Toronto in 1973 and president of UBC in 1985.

Societies join forces to seek NSF cash boost . . .

[WASHINGTON] The three major US societies representing biologists, physicists and chemists combined forces this week to ask Congress for a 10 per cent increase in funding for the National Science Foundation (NSF). Paul Walter, president of the American Chemical Society (ACS), was due to testify before the appropriations subcommittee responsible for NSF on 21 April on behalf of the American Physical Society and the Federation of American Societies for Experimental Biology, as well as the ACS.

The societies want the committee to approve the 10 per cent increase in NSF funding that is in President Bill Clinton's budget request. Their joint testimony — a first for the three societies — is the latest outcome of an effort they began 18 months ago to improve the coordination of their lobbying (see *Nature* 384, 393; 1996).

...as Congress warned over NIH funding

[WASHINGTON] The Clinton administration and the US Congress have been urged not to short-change other sciences as they direct significant budget increases to the National Institutes of Health (NIH). The Ad Hoc Group for Medical Research Funding, a Washington-based advocacy group, in a

proposed 1999 budget being sent to legislators this week, calls for a 15 per cent increase for the NIH this year and for a doubling of the agency's \$13.65 billion budget over five years. But it says that "continued progress in medical research depends upon advancements in related fields" including chemistry, physics, mathematics and engineering. It urges politicians to "consider the breadth of this nation's research efforts as interdependent and to fund them accordingly".

Russia and Japan agree to emissions trade-off

[TOKYO] The Russian president, Boris Yeltsin, and Japan's prime minister, Ryutaro Hashimoto, agreed last week that Japanese companies will help Russian factories and power plants to reduce their emissions of greenhouse gases. In return, Japan will gain credits towards meeting its own emissions reduction targets.

The agreement is believed to be the first such greenhouse gas 'swap' since last year's Kyoto meeting of signatories to the United Nations climate convention, at which the principle of 'joint implementation' was agreed. Japan is struggling to meet its agreed target of reducing emissions by six per cent from 1990 levels during the period 2008 to 2012. In contrast, Russia's closure of inefficient industries means it anticipates little problem in keeping emissions below their 1990 levels.

Science writer Diamond wins Pulitzer book prize



[LONDON] Jared Diamond (left), professor of physiology at the University of California Los Angeles School of Medicine, has won a Pulitzer prize for his book *Guns, Germs and Steel*. The book argues that societies that contributed to

advanced civilizations did so on the basis of geography and environmental factors, not superior ability (see *Nature* 386, 339; 1997).

Guns, Germs and Steel charts the history of agriculture, technology, governments, organized religion and epidemic diseases.

Diamond is an accomplished science writer and a regular contributor to *Nature*. He previously won the 1992 *Los Angeles Times* science book prize for *The Third Chimpanzee*.

Although there is much prestige attached to Pulitzer prizes, which encompass 22 categories of journalism and literature, the prize itself is relatively modest; Diamond will be awarded \$5,000.

Assessing risks of genetic engineering

Sir — Miller *et al.* are right to draw attention to difficulties that might be encountered in considering socioeconomic issues before the release of genetically modified organisms (*Nature* 392, 221; 1998). In fact, it is possible to foresee many more. It may also be the case that “such efforts would constitute major interdisciplinary research projects and would be highly vulnerable to the vagaries of value judgements”, although these are problems that do not preclude productive work in other areas of policy research. (One of Miller *et al.*'s examples, the automobile, is increasingly attracting interdisciplinary scrutiny of precisely the type they eschew.)

However, arguments for the complexity of these issues *per se* can hardly be taken to detract from the importance of efforts that should be made to address them head on. Social and economic factors surely merit overt consideration. We argue, simply, that it is less desirable still to leave them vulnerable to the vagaries of value judgements made tacitly under the constraints of risk assessment frameworks that fail to take account of the need for interdisciplinary input (see *Nature* 391, 528; 1998).

Miller *et al.* claim that consideration of socioeconomic issues would necessarily entail that “any field trial, anywhere... would be subject to an evaluation of possible social and economic impact anywhere else in the world”. With regard to the ‘biosafety protocol’, one option under consideration (article 26) would require the assessment of socioeconomic impact before the transboundary movement of a genetically modified organism. This would not entail that India, for example, need consider the potential economic impact upon Texan farmers of a new variety of genetically modified sorghum, before commencing field trials in Andhra Pradesh. (Whether or not the impact upon farmers in India is considered would of course depend upon Indian

legislation or guidelines.)

Finally, Miller *et al.* allude to the stipulations of free trade agreements. The interaction of the ‘biosafety protocol’ with such agreements is moot. However, there are important precedents. The Montreal protocol (to control the release of ozone-depleting chemicals) is both appropriate and useful, despite its possible conflict with free trade.

Tom Crompton

UNIDO/ICGEB Liaison Office,
Vienna International Centre,
Austria

e-mail: tcrompton@unido.org

John Barrett

Department of Genetics,
University of Cambridge, UK

Chris Gliddon

University of Wales, Bangor, UK

Julian Kinderlerer

Sheffield Institute of Biotechnological
Law and Ethics, University of Sheffield, UK

George Tzotzos

UNIDO Investment and Technology
Promotion Division,
Technology Service, Vienna, Austria

Sir — There seems to be a willingness among many scientists involved in genetic manipulation to speculate favourably at an early stage on the potential economic value of discoveries, with far less enthusiasm for considering potential risks.

Consider this. How often have you seen a grant application or paper in this field that makes a favourable value judgement on the potential value of the work in the concluding paragraph? And how often have you seen similar documents that contain statements giving equal weight to potential biosafety issues or possible adverse economic outcomes? The letter by Miller *et al.* seems to be another manifestation of this mindset, conveying the underlying assumption that there is no justification for attempting to predict wider negative impacts of genetically manipulated organisms

that appear superficially to address pressing environmental, health or food needs.

Miller *et al.*'s scepticism about the value of interdisciplinary research projects aimed at investigating socioeconomic impacts of recombinant organisms during the early phases of field testing overlooks at least one potentially valuable outcome of this work. Such research, predicting the potential impact of changes at the earliest opportunity, would maximize the time available for affected communities to develop contingency measures to buffer the socioeconomic impact of change. Sudden, unpredicted economic changes triggered by the large-scale introduction of genetically manipulated organisms, with no thought for contingencies to mitigate undesirable effects, may be politically and socially destabilizing, particularly in developing countries.

Their argument for dismissing the need for such research — that similar analyses were not required for the introduction of earlier novel technologies such as the internal combustion engine — appears to be breathtakingly complacent and the antithesis of the spirit of scientific enquiry and advancement. Why not extend such logic to safety regulations for all new agrochemicals, and campaign for the retention of standards applied in the nineteenth century? It is perfectly understandable that the genetic engineering community should present an upbeat image of its work to reassure investors, but it is also in their interests to take a longer view of the impact of genetically manipulated organisms. This will help to reassure the public that science can take an active stance on a wide spectrum of biosafety issues.

Phil Gates

Department of Biological Sciences,
University of Durham,
Durham DH1 3LE, UK
e-mail: p.j.gates@durham.ac.uk

Who flew first?

Sir — J. L. Heilbron and W. F. Bynum (*Nature* 391, 13–16; 1998), in “Eighteen-ninety-eight and all that”, celebrate some interesting anniversaries. The choice of such anniversaries is always subjective, because there are so many, but I should like to mention some birth dates that I would have included: 1548 Simon Stevin (Flemish) and Giordano Bruno (Italian); 1598 Bonaventura Cavalieri (Italian); 1698

Thomas Hodgkin (English) and the Frenchmen Charles du Fay, Pierre Bouguer and P. L. Moreau de Maupertuis; 1748 Jean Bernoulli (Swiss) and C. L. Berthollet (French); 1798 Auguste Comte (French).

J. Berzelius (Swedish), B. Bolzano (Czechoslovak) and Caroline Herschel (German) died in 1848, and the Frenchman Marin Mersenne in 1648.

Maybe I can use the authors' bias as an excuse for my own nationalistic and last

comment: how can one mention the Wright brothers as the inventors of the aeroplane without referring to the controversies on this subject? For example, why not a simple allusion to the Brazilian aviation pioneer Alberto Santos Dumont or to the Frenchman Clément Ader?

Ildeu de Castro Moreira

Instituto de Física — UFRJ,
CP 68528, Ilha do Fundão,
21945-170 Rio de Janeiro, Brazil
e-mail: ildeu@if.ufrj.br

Reaping the benefits of basic research

Sir — The benefit of supporting basic research has been increasingly questioned in recent years. Funding agencies in Europe and elsewhere favour applied research, some even appearing to assume that basic research is a luxury that can be done without.

One important argument in favour of basic research, however, is that it consistently yields surprises that in turn are converted into products and even whole new industry sectors, and that, without such continuous innovation, the ensuing stagnation will be damaging to the health and economic strength of society. In the life sciences, the development of monoclonal antibodies and the biotechnology industry are good examples.

But how much basic research is needed to produce the occasional result that goes beyond 'knowledge expansion' and achieves 'usefulness'? The data are sparse, but could be crucial in allowing policy-makers to arrive at more informed decisions. One source of information is the practical consequences of work carried out by scientists who have received grants to do basic research.

The European Molecular Biology Organization (EMBO), which is supported by more than 20 European countries, is well placed to obtain this information, as it has an established postdoctoral fellowship programme where the awards are made solely on the scientific quality of the research proposal. A survey of the career paths of those awarded EMBO fellowships in 1984 and 1985 (see *Nature* **388**, 416; 1997) has been extended to ask whether the research carried out during the fellowship had knowledge as its only outcome, or whether it also led to new products, or ideas for new products.

Of the 120 individuals surveyed, 80 replied, of whom 40% provided examples of practical consequences arising from their basic research. Even making the extreme assumption that the work of all the non-responders resulted only in abstract knowledge, there were still more than 27.5% who could list practical or applied consequences of their work. The list covers all areas of biotechnology, and can be viewed at the EMBO Web site (<http://www.embo.org>). Seven projects gave rise to patents, and many in the 33 examples are of obvious and direct importance in biotechnology today.

Given the outcome of this survey, and the fact that projects that receive EMBO support are judged only on the quality of the science, we should perhaps reassess

policies that attempt to move a significant portion of research into areas justified on the basis of their projected applied outcome. It may be more important to allow decisions on grants to focus on the quality of the research and the researcher, and to ensure that those making the decisions are leaders in the scientific community. Selecting projects because of the promised practical outcome may have the least long-term input to the industries they wish to support.

Frank Gannon

(Executive Director)

EMBO, Meyerhofstr. 1,

D-69117 Heidelberg,

Germany

e-mail: gannon@embl-heidelberg.de

Teaching research skills

Sir — The United Kingdom's experimental programme that teaches students research skills is a step in the right direction¹. The experience may also serve as a model for the United States. Many American graduate schools seem to teach students skills specific to their fields of specialization, but not general principles of research methodology. It is time they considered adding such courses to their graduate programmes.

Scientific research consists of sophisticated intellectual activities that are poorly understood. Rational or logical approaches often fail, whereas irrational methods are sometimes productive. Many scientific discoveries are made by chance rather than as a result of planned research. It is frequently intuition, not rational or conscious thinking, that leads to an innovative theory. This is why many consider the methodology of research an art that cannot be taught. For the same reason, it might be difficult to come up with a method, as desired by Boryeu Mao², that measures fruitless research activities against those that are productive.

Nevertheless, I believe that some general principles of research methodology do exist, and an introduction to such basic knowledge (logic, philosophy of science, skills of thinking and so on) can help young scientists to improve the efficiency of their research. These methods and skills are universal and can be applied to any field. For instance, the reductionist programme has been widely used in research, even though some may not know the term reductionism. It is therefore useful for researchers to understand that this approach is not appropriate for all problems in science^{3,4}. In his time-honoured monograph⁵, W. I. B. Beveridge offered a vivid account of methodology for making scientific discoveries from a

psychological perspective. I recommend this work to science students.

Tianhan Xue

154 Kingston Ave, Suite E,

Santa Barbara, California 93117, USA

1. Masood, E. *Nature* **391**, 729 (1998).
2. Mao, B. *Nature* **391**, 738 (1998).
3. *The Limits of Reductionism in Biology*, Ciba Symposium, Vol. 213 (Wiley, in the press).
4. Agazzi, E. (ed) *The Problem of Reductionism in Science* (Kluwer Academic, Dordrecht, 1991).
5. Beveridge, W. I. B. *The Art of Scientific Investigation* (Heinemann, London, 1961).

Cancer controversy

Sir — Alison Abbott's summary¹ of the controversial cancer therapies of Luigi Di Bella accurately documents the fact that Di Bella has so far resisted efforts to have a committee of experts review his data, which raises further concerns that his combinations of a somatostatin analogue with vitamins and/or melatonin are not as effective as he claims.

The situation is more complex, however, as there is formal preclinical evidence indicating antineoplastic activity of somatostatin analogues such as octreotide and RC-160 in tumour model systems (reviewed in refs 2, 3). These molecules appear to act by lowering systemic IGF-I levels (which may be relevant to cancer risk and prognosis⁴) and/or by activating specific growth inhibitory signal transduction pathways linked to specific somatostatin receptors which are expressed by neoplastic cells⁵. The preclinical evidence is sufficiently impressive, particularly in low tumour burden models, for both the US National Cancer Institute and the National Cancer Institute of Canada to be conducting randomized clinical trials to evaluate the activity of octreotide in the adjuvant treatment of breast cancer. A trial based at the Mayo Clinic comparing tamoxifen to the combination of tamoxifen and low dose octreotide in the treatment of advanced breast cancer showed no difference between treatment arms⁵.

Somatostatin analogues may or may not be shown in formal clinical trials to be useful in the treatment of specific neoplastic diseases, but it would be a mistake to use Di Bella's conduct to discredit rigorous research in this area.

Michael Pollak

Department of Oncology,

Jewish General Hospital-McGill University,

Montreal, Quebec, Canada H3T 1E2

e-mail: MD49@musica.mcgill.ca

1. Abbott, A. *Nature* **391**, 217 (1998).
2. Weckbecker, G. et al. *Cancer Res.* **54**, 6334 (1994).
3. Pollak, M. & Schally, A. V. *Proc. Soc. Exp. Biol. Med.* **217**, 143 (1998).
4. Chan, J. et al. *Science* **279**, 563 (1998).
5. Ingle, J., Kardinal, C., Suman, V., Krook, J., Pollak, M. *Proc. Amer. Soc. Clin. Oncol.* **16**, 150 (abstr. 526).

Ending up with the right partner

Titia de Lange

How do the correct maternal and paternal chromosomes find one another during sexual reproduction? New data from fission yeast indicate that the search for the right partner is simplified by pre-alignment of the chromosomes through clustering of their ends.

Arguably the most intimate event in sexual reproduction is the tight embrace of paternal and maternal chromosomes that occurs during the formation of gametes (in humans, the eggs or sperm). In a process known as meiosis, the two copies of each chromosome (called homologues) snuggle up to form precisely aligned pairs and exchange genetic information (recombination). The pairing and recombination of homologues permits their

faithful separation, such that each gamete is haploid — that is, endowed with just one copy of each chromosome. But how do homologous chromosomes meet? On pages 825 and 828 of this issue, Nimmo *et al.*¹ and Cooper *et al.*² pinpoint the ends of chromosomes (telomeres) as major movers in homologous pairing in the fission yeast *Schizosaccharomyces pombe*.

The problem of homologue pairing is one of staggering complexity. During meiosis in

the human germ line, for instance, the search for homology takes place amongst the more than 6 million kilobase-pairs of DNA organized on 23 chromosome pairs. Early cytologists noted that, in a phase of meiosis called prophase, chromosomes tend to be positioned with their telomeres clustered at one pole of the nucleus and the chromosomal arms looping out in the other direction. Known as the bouquet stage (see Fig. 1, and ref. 3 for review), this configuration might promote chromosome pairing. However, it has been hard to progress from this descriptive work to testing experimentally how telomeres are involved in meiosis.

Fission yeast has emerged as an invaluable tool to address this question. Unlike most eukaryotes, this organism prefers a haploid lifestyle, opting for meiosis as soon as the joining of two haploid cells has generated the diploid zygote. The formation of a zygote in fission yeast involves a specialized kind of nuclear fusion (karyogamy) that unites the two haploid nuclei of the parental cells (Fig. 1). Subsequent meiotic events include DNA replication, a prolonged prophase and, finally, the two meiotic divisions that segregate the homologues and give rise to four haploid spores.

An unusual feature of *S. pombe* meiosis is that its newly formed zygotic nucleus is highly mobile in prophase I, moving back and forth through the cell during the exact period in which homologues pair and recombine⁴. These striking nuclear oscillations have been captured by video microscopy, and they could explain the elongated 'horsetail' shape of prophase I nuclei that had previously been noted in fixed zygotic cells⁴. The horsetail movement is driven by a structure called the spindle pole body (SPB), which is embedded in the nuclear envelope and connects to dynamic microtubules.

Where do the telomeres come in? Normally, the SPB is associated with a specialized region of the chromosome known as the centromere. But, at the horsetail stage, telomeres cluster at this position^{4,5}. This centromere–telomere switch occurs in haploid cells when they are exposed to mating pheromone⁵. Thus, telomeres end up at the leading edge of the elongated nucleus, with the rest of the chromosomes trailing behind as the SPB moves the nucleus back and forth through the cell. This curious mechanism seems particularly well-suited to orientate the chromosomes and position homologous sequences close together. In effect, it is similar to grabbing a bunch of ropes by their ends and giving them a good shake — the ropes will align according to their lengths. Because the *S. pombe* genome is organized on just three chromosomes of different sizes, it has been proposed that the search for homology in meiosis is facilitated considerably by arranging the chromosomes as loops held together by their ends.

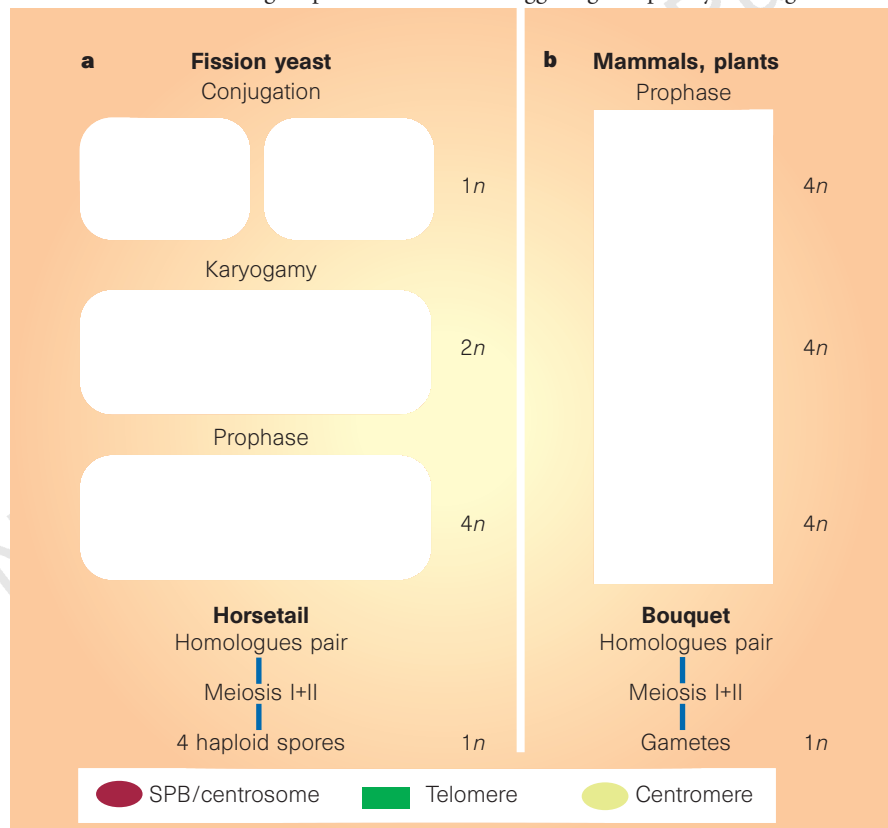


Figure 1 Chromosome movements that occur before pairing of maternal and paternal (homologous) chromosomes during sexual reproduction in fission yeast, mammals and plants. **a**, Fission yeast. After cellular joining (conjugation) and nuclear fusion (karyogamy) between two haploid ($1n$) cells with opposite mating type, a zygote ($2n$) is formed. Meiosis (the formation of gametes) is initiated immediately and, in a stage of meiosis I known as prophase, nuclei undergo rapid movements back and forth through the cell, led by the spindle pole body (SPB). The ends of the chromosomes (telomeres) are connected to the SPB, and the telomere-led movement of the so-called 'horsetail' nucleus facilitates alignment of homologous chromosomes. Nimmo *et al.*¹ and Cooper *et al.*² have shown that loss of the telomere–SPB connection results in reduced exchange of genetic information and diminished spore viability. **b**, Mammals and plants. Telomeres similarly cluster at one pole of the nucleus (in mammals, at a structure called the centrosome, which is analogous to the SPB) in prophase of meiosis I. The resulting configuration of the chromosome is known as the bouquet stage.

Nimmo *et al.*¹ and Cooper *et al.*² now show that this view is likely to be correct, and that telomeres are important in the pairing and recombination of homologues. Their findings arose from an interest in the proteins that form a complex at the telomere. The product of the *taz1* gene, Taz1p, was identified⁶ in Tom Cech's lab from a one-hybrid screen for factors that bind telomeric DNA. It was shown to affect telomeric silencing and the maintenance of telomere length. The findings also hinted at a meiotic function for Taz1p, because the spore viability was greatly reduced in the mutant strain⁶.

Nimmo *et al.*¹ carried out a genetic hunt for genes involved in telomere function, using telomeric silencing as a screening tool. They identified the *lot2* gene (as well as *taz1*), and showed that a mutation in *lot2* (like *taz1*) results in inappropriate telomere elongation and a marked reduction of viable spores.

The nature of the meiotic defect in *taz1* and *lot2* mutants became clear when the positioning of telomeres at the SPB was examined — rather than showing the normal clustering of all chromosome ends at the SPB, both Nimmo *et al.* and Cooper *et al.* found that most of the telomeres were scattered in the mutants.

Nimmo *et al.* recorded the behaviour of defective telomeres, and their effect on the horsetail stage, by video microscopy. They had previously found that a protein known as Swi6p localizes to both centromeres and telomeres, where it is involved in controlling transcriptional repression⁷. They now use a fusion protein between Swi6p and green fluorescent protein to label telomeres and centromeres as bright fluorescent dots in living cells. The authors produced a movie⁸ documenting the vigorous swinging back and forth of the wild-type horsetail nucleus, with telomeric Swi6p dots in the lead and centromeric dots lagging behind. In the *lot2* mutant, however, the telomeric dots fail to cluster at the SPB, and are dispersed throughout the nucleus. The SPB still seems to try and move the nucleus through the cell but, because there is no connection to the chromosomes, it merely manages to drag a small bit of nuclear material back and forth.

Loss of the connection between telomeres and the SPB — and the resulting defect in movements at the horsetail stage — has severe consequences for recombination between homologous chromosomes. The *taz1* and *lot2* mutants showed^{1,2} a marked reduction (3–10-fold) in the rate of recombination both within and between genes. This defect could be even more extreme than these figures imply, because over 90% of the mutant spores were not viable — probably because the chromosomes had failed to segregate normally in meiosis I. These data indicate that telomere clustering at the SPB is required for the recombination and segregation of homologous chromosomes, in

keeping with the idea that the telomere-led horsetail movement facilitates chromosome pairing.

What is the significance of these findings with regard to meiosis in other organisms? It is now clear that meiotic strategies can be quite diverse⁹, and a horsetail stage in strictest sense has not been seen in budding yeast, fruit flies or humans. But the meiotic tango of fission-yeast chromosomes is likely to be more than a cell-biological curiosity. After all, mammalian and plant chromosomes also move — when their telomeres adhere to the nuclear envelope in prophase and then gather at one pole of the nucleus to form the bouquet stage^{10,11}. Moreover, the timing of the bouquet stage is compatible with its role in the pairing of homologous chromosomes, because telomeric clustering precedes alignment of the homologues and formation of the synaptonemal complex, an elaborate structure that holds the homologues together until recombination has been completed¹¹.

A further hint for a telomere-mediated pairing mechanism in mammals is the functional and structural similarity of Taz1p to the mammalian telomeric protein TRF1 (refs 6, 12, 13). TRF1 can promote parallel

pairing of telomeric tracts *in vitro*¹⁴, an activity that is potentially significant to meiosis. Mice lacking TRF1 should reveal the function of this protein in the formation of mammalian gametes. Human telomeres have been implicated in cellular ageing and in cancer — sex may be next. □

Titia de Lange is at The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

e-mail: delange@rockvax.rockefeller.edu

1. Nimmo, E. R., Pidoux, A. L., Perry, P. E. & Allshire, R. C. *Nature* **392**, 825–828 (1998).
2. Cooper, J. P., Watanabe, Y. & Nurse, P. *Nature* **392**, 828–831 (1998).
3. Dernburg, A. F., Sedat, J. W., Cande, W. Z. & Bass, H. W. in *Telomeres* (eds Blackburn, E. H. & Greider, C. W.) 295–338 (Cold Spring Harbor Laboratory Press, New York, 1995).
4. Chikashige, Y. *et al. Science* **264**, 270–273 (1994).
5. Chikashige, Y. *et al. EMBO J.* **16**, 193–202 (1997).
6. Cooper, J. P., Nimmo, E. R., Allshire, R. C. & Cech, T. R. *Nature* **385**, 744–747 (1997).
7. Ekwall, K. *et al. Science* **269**, 1429–1431 (1995).
8. <http://www.nature.com>
9. McKim, K. S. *et al. Science* **279**, 876–878 (1998).
10. Bass, H. W., Marshall, W. F., Sedat, J. W., Agard, D. A. & Cande, W. Z. *J. Cell Biol.* **137**, 5–18 (1998).
11. Scherthan, H. *et al. J. Cell Biol.* **134**, 1109–1125 (1996).
12. Chong, L. *et al. Science* **270**, 1663–1667 (1995).
13. van Steensel, B. & de Lange, T. *Nature* **385**, 740–743 (1997).
14. Griffith, J., Bianchi, A. & de Lange, T. *J. Mol. Biol.* (in the press).

Planetary science

Comet leaves a trail in the air

Laura Garwin

The inner Solar System is a dusty place. As the Earth makes its way around the Sun, it travels through a tenuous cloud of dust, the 'zodiacal cloud', thought to be produced mainly by colliding asteroids and active comets, with a small contribution of dust from outside the Solar System. About 10,000 tonnes of this dust hits the Earth each year, and a tiny fraction finds its way — with human aid — into laboratories, under microscopes and through mass spectrometers. The compositions of these interplanetary dust particles (IDPs) show that they are extraterrestrial, but a more detailed understanding of their origin, such as whether a given particle has come from an asteroid or comet, has been elusive. Now some of these particles have been coaxed into giving up a bit more of their life story, according to a talk given last month*. Scott Messenger (Nat'l Inst. Standards Technol., Gaithersburg) suggested that a population of the fluffy, fragile aggregates known as 'cluster IDPs' collected in the stratosphere in the summer of 1991 come not just from a comet, but specifically from the periodic Earth-crossing comet Schwassmann–Wachmann 3 (Fig. 1). If Messenger is right, dust from this comet, and perhaps from several others, can be sampled

repeatedly using nothing more sophisticated than a high-altitude aircraft.

The large interplanetary dust particles known as cosmic spherules can be collected from the sea floor and from polar ice. But melting on atmospheric entry and subsequent alteration on Earth lead to changes in composition that limit their usefulness as samples of extraterrestrial material. The place to find nearly pristine IDPs is the stratosphere. Since 1974, NASA has flown decommissioned U-2 spy planes at about 20 km altitude, with simple impact collectors mounted under the wings. These planes are usually 'ships of opportunity', collecting dust while performing other research, so it can take months for a collector to accumulate the 30–40 hours of exposure required to harvest a useful crop of a few hundred IDPs, ranging from about 5 to 50 µm across. IDPs are identified as extraterrestrial by the presence of solar-flare particle tracks, implanted noble-gas atoms from the solar wind, and by non-terrestrial isotope ratios of hydrogen, nitrogen and oxygen. Most IDPs also have major-element compositions that are similar to those of the primitive meteorites known as carbonaceous chondrites, consistent with an origin in comets or certain types of asteroid. Specifically, the more porous IDPs are thought to come from comets¹. But without

*29th Lunar and Planetary Science Conference, Houston, Texas, 16–20 March, 1998.

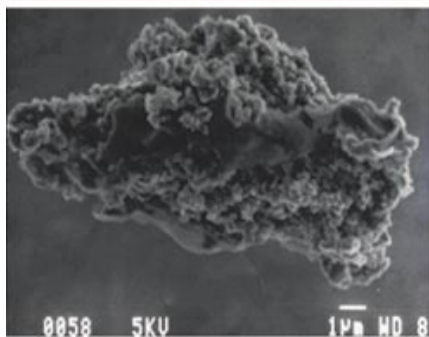
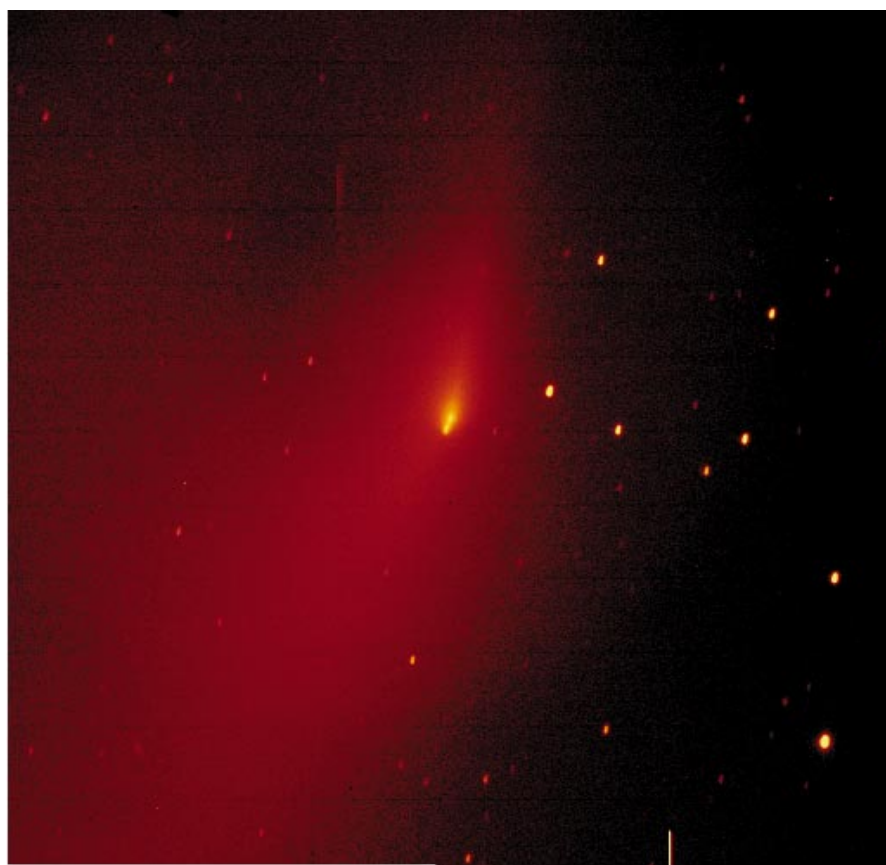


Figure 1 Mother and child? The comet Schwassmann–Wachmann 3, and (left) an interplanetary dust particle captured by a high-flying aircraft. A set of these particles has been tentatively traced to Schwassmann–Wachmann 3, and so may be the nearest thing to pristine comet material that we have.

any samples of certified comet dust for comparison (or samples of asteroids, other than the highly biased selection provided by meteorites), the connection has been hard to prove.

Enter Messenger. He noticed that cluster IDPs — so called because on striking the dust collector, they break up into dozens or hundreds of tiny fragments — typically have more extreme and more variable hydrogen and nitrogen isotope compositions² than individual IDPs. Moreover, this is most clearly displayed by cluster particles collected in June and July 1991 (JJ-91) — a population that had already been identified as “unrepresentative of those generally collected”³, on the basis of their extraordinarily low helium contents and high ³He/⁴He ratios.

Messenger realized that the low helium content of the JJ-91 particles can be used to constrain the time that the particles spent in space, travelling between their parent body and the Earth’s atmosphere. It takes only a few years for a dust particle at the Earth’s dis-

tance from the Sun (1 AU) to become saturated with helium from the solar wind. As the JJ-91 particles are not saturated, they can have been exposed in space for no more than about ten years. This in itself is a striking conclusion, as the average lifetime of dust near 1 AU is about 10,000 years.

This extremely short exposure age puts powerful constraints on the source of the JJ-91 dust⁴. On such a short timescale, only radiation pressure from the Sun can modify the dust particles’ orbits relative to that of their parent body, and as radiation pressure will always enlarge the orbits, the parent body must approach closer than 1 AU to the Sun.

As the Earth-crossing asteroids do not produce much dust, Messenger argues that the field of possible parent bodies is restricted to 17 currently active Earth-crossing comets. Of these, only four have orbits with eccentricity low enough to account for the low peak temperatures experienced by the JJ-91 particles on entering the atmosphere. (A higher eccentricity would have given the par-

ticles a higher entry velocity, leading to more loss of volatiles than is observed.) And of these four, only one comet — the aforementioned Schwassmann–Wachmann 3 — crosses the Earth’s orbit at the right time of year to have produced particles collected in June and July.

As with some of the best new ideas, Messenger’s hypothesis seems both obvious and impossible. Meteor showers mark the passage of the Earth through the dusty trails of comets, so collecting the dust from a specific comet should be an obvious idea. But only recently have collection periods as short as 2–3 months become routine, allowing the time resolution necessary to spot contributions from a specific comet. On the impossible side of the balance sheet, it seems that we must be extraordinarily lucky to capture dust within ten years of its release from a comet, when the lifetime of an average dust particle is more like 10,000 years. But the hypothesis can be easily tested by making repeated collections at the same time of year, and comparing the nature and amount of dust collected with the observed activity and location of Schwassmann–Wachmann 3.

And what if Messenger is right? With nearly pristine samples of a comet in hand, it should be possible to attack long-standing questions of comet origin and composition, such as whether comets can contain hydrous minerals, and how well they preserve material from before the birth of the solar nebula. Short-timespan collections at other times of year could be made to look for dust from the three other low-eccentricity Earth-crossing comets: Grigg–Skjellerup, Machholz and Wilson–Harrington. Ground- and space-based observations of comets have already shown that they can be very different from one another, so samples from many comets will be necessary for any complete understanding of comet formation and evolution.

Convenient as it is to collect comet dust without leaving home, the dust collected in the stratosphere is not truly pristine; its years in space, weeks in the atmosphere, and final violent encounter with the silicone-oil-coated collector will have had some effect on its composition and structure. For this reason, both comet and interplanetary dust specialists will still be looking forward to the STARDUST mission⁵, which in 2006 is scheduled to return to Earth dust and volatiles from comet Wild 2, gently captured in aerogel collectors. □

Laura Garwin is the North American editor of *Nature*.

e-mail: l.garwin@nature.com

1. Brownlee, D. E. *Annu. Rev. Earth Planet. Sci.* **13**, 147–173 (1985).
2. Messenger, S. & Walker, R. M. in *Astrophysical Implications of the Laboratory Study of Presolar Materials* (AIP conference vol. 402; in the press).
3. Nier, A. O. & Schlutter, D. J. *Meteoritics* **28**, 675–681 (1993).
4. Messenger, S. & Walker, R. M. <http://cass.jsc.nasa.gov/meetings/LPSC98/pdf/1906.pdf>
5. <http://stardust.jpl.nasa.gov/>

Biological sensors

Controlling the fly's gyroscopes

Roland Hengstenberg

True flies — such as hoverflies or the blow fly *Calliphora vicina* — have breathtaking aerobatic capabilities due to a very elaborate flight apparatus. Not only do they beat their wings up to 150 times per second, but they have a gearbox with three gears in the wing joint¹, and use non-stationary aerodynamics to generate exceptionally large flight forces². These flies also show masterly control as they fly around obstacles and through turbulent air, and the way in which they do this is revealed in part by Chan *et al.*³ in *Science*. The authors report unexpected features of the 'gyroscopic' sense organs that tell the fly about its rotations in space.

Most insects have four wings, consisting of the thin wing blade spread between a framework of stiff veins that also carry touch and strain receptors⁴. Wings are, therefore, usually mechanical effectors and sense organs at the same time. The wings of true flies are driven by two kinds of specialized muscle. First, two large 'power' muscles (which fill most of the fly's thorax) contract antagonistically at wing-beat frequency, in mechanical resonance with the thoracic box⁵. They produce several contraction cycles for each impulse from a motor neuron, so they are also called 'asynchronous' muscles⁵. Second, 13 small 'control' muscles, operating synchronously, act on the wing joint to extend and retract the wings, and to modify the wing kinematics for steering⁵.

During evolution, the hind wings from the ancestors of true flies were transformed into the so-called halteres^{6–8}. These are small, club-shaped organs, buried in the cleft between the fly's thorax and abdomen (Fig. 1a). During flight they oscillate up and down around a hinge joint (Fig. 1b), through an angle of about 180° and in antiphase with the wings. Because of their small size and shape, as well as their location, they can hardly be expected to have any aerodynamic effect. Instead, they have been shown to act as highly specialized 'gyroscopic' sense organs, measuring rotations of the fly in space^{6–8}.

An oscillating haltere tends to preserve its angular momentum. When the fly turns, a periodic Coriolis force is generated perpendicular to the plane of the oscillation. The Coriolis force means that a mass moving in a rotating system is accelerated perpendicular to its motion and to the axis of rotation. This force acts mainly on the haltere knob (Fig. 1b), and its amplitude is proportional to the vector product of the haltere's linear velocity and the fly's angular velocity. So, halteres sense angular velocity directly — unlike the vertebrate semicircular canals, which integrate angular acceleration.

The frequency of the Coriolis force depends on the direction in which the fly is turning. For pitch turns around the transverse axis, the force has the same frequency as the wing beat. But for yaw turns around the vertical axis, the force has twice the wing-beat frequency^{6,8}. Moreover, although a single haltere is 'blind' for rotations around the axis of its oscillation⁹, roll turns of the fly around the longitudinal body axis can still be sensed because the two halteres are arranged obliquely (Fig. 1a). The necessary distinction of pitch and roll turns is achieved by bilateral processing of the signals from the two halteres^{6–10}, probably in the ventral cord (the insects' analogue of the spinal cord).

The reactive force acting on the centre of mass — that is, on the haltere knob — is transmitted along the stiff stalk (Fig. 1b) to the base of the haltere, where it produces a very complicated spatio-temporal strain pattern in the insect's cuticle. This pattern is encoded by about 335 cuticular strain receptors, most of which are clustered in five neatly ordered fields around the base of the haltere (Fig. 1b)¹¹. The central nervous pathways that process the signals from these mechanoreceptors are poorly understood. But behavioural studies have shown that flies use the information from their halteres to stabilize the orientation of their head in space^{7–10}, to maintain flight equilibrium¹² and to follow their intended course¹³.

Because *Calliphora* oscillates its halteres as it walks — when the wings are held perfectly still — each haltere must be moved by its own set of muscles. Until now, this has been thought to involve either one or two

power muscles, like those that control the wings. My own observations show that there are two, a depressor and a levator, as proposed by Schneider¹⁴. But Chan *et al.*³ claim that there is only one power muscle, as proposed 50 years ago by Pringle⁶. Furthermore, the authors describe a set of 11 tiny control muscles, which had never been seen before, at the base of the haltere. Eight of them originate on the hard body wall, and are inserted at various points on minute sclerites of the haltere joint. From these sites of attachment, the authors infer that each is homologous to the corresponding muscle of the wing joint.

The mechanosensory function of the halteres therefore seems to be adjustable by commands from the fly's nervous system. This might involve adaptations to different usage of the halteres when either walking or flying, or it may be a sign of fine tuning of the haltere kinematics. Or, as proposed by Chan *et al.*, it may be a means of particularly efficient flight control, to modulate signals from the halteres that converge on some of the flight control muscles.

The most surprising new finding is that at least two of the eight control muscles receive input from the visual system. Chan *et al.* recorded (extracellularly) the spike activity of the control muscles B2 and I1 in resting flies, while moving striped patterns were presented in the frontal visual field of the experimental fly. They found that the impulse activity of the muscles was modulated by the pattern motion in a directionally specific manner. For example, whereas B2 was maximally excited when the pattern moved downwards and a little to the side of the recorded muscle, I1 responded best to motion 45° upwards and to the same side. The significance of these directional specificities is not yet clear, and the different effects of visual inputs on the two mechanosensory axes of the haltere still have

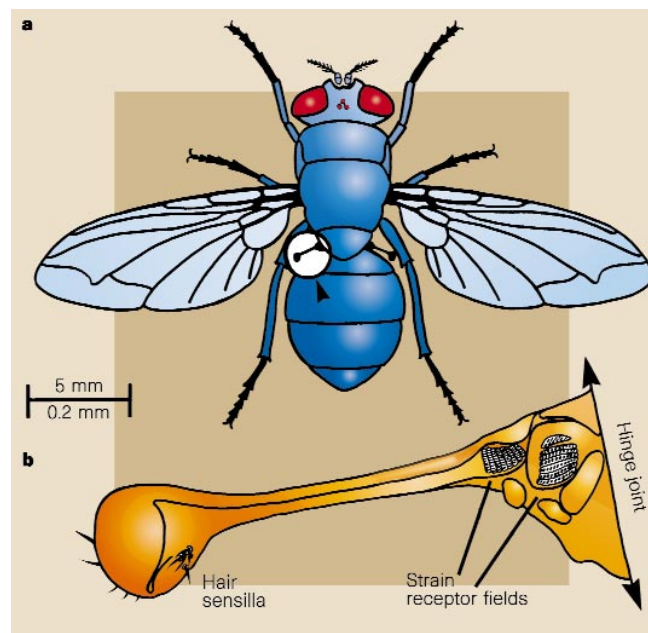


Figure 1 Halteres, the 'gyroscopic' sense organs of the blow fly *Calliphora vicina*. During walking and flight, the halteres oscillate in a vertical plane around a proximal hinge, and Chan *et al.*³ have now shown that they are under both neural and visual control. a, The halteres (arrowhead) are found between the thorax and abdomen of a fly. b, Left haltere from above. Most of the mass of the haltere is in the knob (left). A thin, stiff stalk leads to a base that houses about 335 cuticular strain receptors.

to be worked out. But, because halteres sense rotatory self-motions, it seems reasonable to assume that the observed visual motion sensitivities must be interpreted in this context. Thus, although the halteres have been thought of as an extremely specialized, hard-wired and purely mechanosensory system, Chan *et al.*³ have forced us to reconsider the fly's 'gyroscopes' as sense organs that are under neural and visual control. □

Roland Hengstenberg is at the Max-Planck-Institut für Biologische Kybernetik, Spemannstrasse 38, D-72076 Tübingen, Germany, and at the Institute for Advanced Study, Wallotstrasse 19, D-14193 Berlin, Germany.

e-mail: roland.hengstenberg@tuebingen.mpg.de

1. Nalbach, G. *J. Comp. Physiol.* **165**, 321–331 (1989).
2. Lehmann, F.-O. & Dickinson, W. H. *J. Exp. Biol.* **200**, 1133–1144 (1997).
3. Chan, W. P., Prete, F. & Dickinson, W. H. *Science* **280**, 289–292 (1998).
4. Dickinson, M. H. *J. Exp. Biol.* **169**, 221–233 (1992).
5. Dickinson, M. H. & Tu, M. S. *Comp. Biochem. Physiol.* **116A**, 223–238 (1997).
6. Pringle, J. W. S. *Phil. Trans. R. Soc. Lond. B* **233**, 347–384 (1948).
7. Hengstenberg, R. *J. Comp. Physiol.* **163**, 151–165 (1988).
8. Nalbach, G. *J. Comp. Physiol.* **173**, 293–300 (1993).
9. Nalbach, G. *Neuroscience* **61**, 149–163 (1994).
10. Nalbach, G. & Hengstenberg, R. *J. Comp. Physiol.* **175**, 695–708 (1994).
11. Gnatzky, W., Grünert, U. & Bender, M. *Zoomorphology* **106**, 312–319 (1987).
12. Faust, R. *Zool. Jahrb. Physiol.* **63**, 325–366 (1952).
13. Heide, G. in *Insect Flight* (ed. Nachtigall, W.) 35–52 (G. Fischer, Stuttgart, 1983).
14. Schneider, G. Z. *Vergl. Physiol.* **35**, 416–458 (1953).

The reconstruction is based on large-scale patterns of temperature fluctuations in the twentieth century as recorded by instruments. The authors derived an empirical relationship between the strength of these patterns in the twentieth century and the palaeoclimate indices, which were then used to estimate the magnitude of the same patterns in times before instruments were used to record temperature. This approach differs from that taken in producing local temperature reconstructions, on which other reconstructions of global-scale temperature are based (see, for example, refs 2 and 3), and enables use of instrumental data other than temperature and of proxy data influenced by more than temperature. For example, tree-ring data may reflect, among other influences, a combination of temperature and precipitation. Mann and colleagues attempt to use the relation of such variables with atmospheric dynamics to reconstruct large-scale temperature patterns.

Given the novelty of this approach, it is not surprising that the uncertainties need more investigation⁴. As the authors acknowledge, climate reconstructions can only be as good as the underlying data. The value of different data as temperature proxies³ varies, which should also influence the pattern reconstructions. Similarly, problems in dating some records may decrease the quality of the reconstruction. Also, we need to know more about the influence of non-resolved patterns on global-scale averages, about the validity of the assumption that proxy and temperature data are related linearly, and about the quality of the reconstruction in the earlier part of the record where data are sparser.

Still, when Mann and colleagues' reconstruction is compared with other new recon-

Climate change

The past as guide to the future

Gabriele Hegerl

We know that in recent years the Earth's climate has been getting warmer. But is this warming unusual relative to low-frequency variations in temperature? If it is, then how much of it has been caused by anthropogenic changes in the composition of the atmosphere?

On page 779 of this issue¹, Mann *et al.* provide a reconstruction of past temperature back to the year AD 1400, extending previous records further into the past, and deriving annual temperature anomaly patterns over large parts of the globe. As examples, the authors have produced reconstructions of the 'year without summer' (1816), which was probably influenced by the eruption of Mount Tambora in Indonesia, and of 1791 when a strong El Niño is known to have occurred. Such reconstructions are poten-

tially useful for interpreting the warming trend in the twentieth century. For example, the reconstructed record for the Northern Hemisphere suggests that the warming is unprecedented, at least since 1400. Also, from a multivariate correlation between the reconstructed temperature time-series for the Northern Hemisphere and external climate influences, it seems that increases in greenhouse gases have been the main forcing in the twentieth century, whereas natural climate forcing by changes in solar irradiance and volcanism dominate the early part of the record.

Mann *et al.* use a quite original and promising method to produce their reconstructions, employing data from tree rings, ice cores, ice melt indices and long historical records of temperature and precipitation.

100 YEARS AGO from 21 April 1898.



50 YEARS AGO from 24 April 1948.

NOTTINGHAM AND DISTRICT
TECHNICAL COLLEGE
Shakespeare Street, Nottingham
Principal:

H. BAKER, Ph.D., A.M.I.C.E., A.M.I.M.E.

Applications are invited for the post of Senior Assistant Lecturer in the following subjects: dyeing and finishing, mainly with reference to the hosiery trade. Hosiery manufacture.

Applications are also invited for the following posts: Responsible Lecturer in Organic Chemistry. Lecturer in Physics, preferably with subsidiary chemistry.

The salary scales are those of the Burnham (Technical) Report. For Senior Assistants the scale ranges from a minimum of £700–£790 to a maximum of £800–£905 (man) or from £560–£632 to £640–£724 (woman). For the Lecturer (if a man) the scale ranges from a minimum of £300–£390 to a maximum of £555–£660; the commencing salary may in certain cases be increased by an amount up to £180 on account of industrial or teaching experience. The salary of the Responsible Lecturer carries an additional allowance of £75 per annum.

Further particulars and form of application may be obtained from the Registrar of the College, to whom the completed application should be returned within a fortnight of the appearance of this advertisement.

F. STEPHENSON,
Acting Clerk to the Governors.

CADBURY'S

COCOA is COCOA,

and COCOA ONLY.

CADBURY'S is not a combination of drugs or a high-sounding
alkalised article, it is **A** most sustaining and
PERFECT
refreshing beverage, being a **FOOD** and a drink combined.

It is specially suitable **FOR** all times and all seasons.

It is most **YOUNG** building up the
nourishing for the growing body.

For breakfast, supper, **AND** all times.

The **OLD** repairing the waste
ideal drink for the incidental to life.

CADBURY'S {ABSOLUTELY
PURE} **COCOA.**

"WE ARE GLAD TO GIVE
IT THE HIGHEST
PRAISE."
—Medical
Magazine.

The Analyst says—
"CADBURY'S is
the TYPICAL
COCOA of
English
Manufac-
ture."

The
Lancet
says—
"CADBURY'S
represents the
standard of
HIGHEST PURITY."

The Medical
Magazine says—
"CADBURY'S,
the most popular
Cocoa of the day."

structions, as described elsewhere⁴, the main features are broadly the same. Taking the record as a whole, most of the twentieth century has been unprecedentedly warm, whereas the nineteenth and the seventeenth centuries were quite cool. The comparison, however, also reveals quite substantial differences in the estimated magnitudes of the warm and cold phases, meaning that work still needs to be done to improve reconstructions of past temperature averages.

In the light of these uncertainties, Mann and colleagues' interpretation of the climate record — in terms of correlations with different forcing agents (greenhouse gases, solar radiation changes and volcanism) — should be considered primarily as a consistency check on results from more rigorous 'fingerprint detection' techniques^{5–7}. In this line of research, the model-simulated evolution of the climate response to anthropogenic climate change, in both space and time, is used to detect such a change in the observations and distinguish it from internal climate variations and from the response to the naturally occurring forcing mechanisms. Thus far, the evidence from these fingerprint detection studies suggests that a significant part of climate change is indeed attributable to human activity⁸. Mann and colleagues' method uses time information only, and assumes a linear response of climate to forcing with no time lags. But their results support, independently of climate models, the conclusion that anthropogenic influences have dominated the evolution of temperature in the twentieth century.

Our incomplete understanding of natural climate variability is the largest limitation in fingerprint detection work. In most cases the internal variability of atmosphere–ocean general circulation models is used as a substitute for instrumental observations, which cover too short a time span to sample natural climate variability on the relevant timescale of several decades. Also, model simulations are used to estimate the influence of natural external forcings on climate (for example ref. 9).

A very useful application of reconstructions of past temperature variability would therefore be to verify model estimates of natural climate fluctuations¹⁰, and of the climate's response to past changes in solar radiation and volcanism. The paper of Mann *et al.* is an important step towards reconstructing space–time records of historical temperature patterns, which could reduce the gap in our knowledge of that variability. Such work can not only provide a window on the past but it can also deliver invaluable information for a more reliable detection of anthropogenic climate change. This in turn brings us closer to glimpsing what the future may hold for a world with increasing levels of greenhouse gases. □

Gabriele Hegerl is at the Joint Institute for the Study of the Atmosphere and Ocean, University of

Washington, Seattle, Washington 98195-4235, USA.
e-mail: hegerl@atmos.washington.edu

1. Mann, M. E., Bradley, R. S. & Hughes, M. K. *Nature* **392**, 779–787 (1998).
2. Bradley, R. S. & Jones, P. D. *Holocene* **3**, 367–376 (1993).
3. Jones, P. D., Briffa, K. R., Barnett, T. P. & Tett, S. F. B. *Holocene* (in the press).
4. Jones, P. D. *Science* **280**, 544 (1998).
5. Santer, B. D. *et al.* *Nature* **382**, 39–46 (1996).

6. Hegerl, G. C. *et al.* *Clim. Dyn.* **13**, 613–634 (1997).
7. North, G. R. & Stevens, M. J. *J. Clim.* (in the press).
8. Santer, B. D., Wigley, T. M. L., Barnett, T. P. & Anyamba, E. *The IPCC Second Scientific Assessment* (eds Houghton, J. T. *et al.*) 407–444 (Cambridge Univ. Press, 1996).
9. Cubasch, U., Voss, R., Hegerl, G. C., Waszkewitz, J. & Crowley, T. J. *Clim. Dyn.* **13**, 757–767 (1997).
10. Barnett, T. P., Santer, B. D., Jones, P. D., Bradley, R. S. & Briffa, K. R. *Holocene* **6**, 255–263 (1996).

Animal behaviour

The evolution of menopause

Paul W. Sherman

Menopause is the permanent cessation of menstruation, and it affects all women by the time they are about 50 years old. As well as raising personal, medical and social issues, menopause is an evolutionary puzzle — because natural selection favours increased reproduction, how could terminating it early be beneficial? On page 807 of this issue, Packer *et al.*¹ explore this question, using data from free-living olive baboons and lions (Fig. 1). They conclude that menopause is a non-adaptive result of senescence (ageing).

Four hypotheses have been proposed to explain menopause. Two are non-adaptive and two adaptive. First, it may simply be a cultural artefact. According to this 'blessings of modern life' theory², menopause occurs because women live longer now than in the past. Most animals reproduce as long as they live, but zoo specimens and pets (whose lifespans have been artificially lengthened) often stop reproducing before they die³. Menopause may similarly result from medically increasing the lifespan of a primate that is born with a fixed total number of gametes⁴.

Second, menopause may be explained by 'senescence' theory. Senescence occurs because the product of reproductive value

and survivorship of young individuals always exceeds that of old individuals⁵. Tendencies to be vigorous, healthy and fertile pile up in youth, whereas somatic and reproductive maintenance is neglected later in life. Thus, menopause may be like geriatric memory loss and failing eyesight — the uncontrolled deterioration of a physiological process that once was well regulated.

But these non-adaptive hypotheses leave some questions unanswered. Why does gamete production terminate abruptly in women, yet taper off gradually in men⁶? And why does ovulation cease so early in a woman's life? Although average life expectancy has increased due to medical advances, this results mainly from reduced juvenile mortality⁷, and the maximum human lifespan (about 115 years) has changed little for centuries. Even in traditional, hunter–gatherer societies, many women live long enough to experience menopause⁸. Moreover, the complex, seemingly programmed way in which menopause occurs suggests adaptation rather than pathology^{6,7,9}.

The first adaptive theory proposes that, because birth defects increase with maternal age, menopause protects the human gene pool. The flaw in this 'group selection' theory



Figure 1 Good mothers — lioness and female olive baboon with young. Packer *et al.*¹ have studied these animals and conclude that reproductive cessation is a non-adaptive consequence of senescence, rather than an adaptation to enable grandmothers to help rear their daughters' young.

BERNARD CASTELEIN

GERRY ELLIS/BBC NAT. HIS.

is that "... if pollution of the gene pool were all that was involved, genes that tended to cause women to continue having babies despite the risk involved would certainly outreproduce those that influenced women to stop having babies"⁹.

Alternatively, menopause may have evolved as a counter-strategy to senescence. According to this 'good mother' theory, "... if the mother has more babies (with associated dangers) even as the ravages of age become severe, she is having children she may not be able to care for, and she is risking the future success of her existing children. If, instead, she stops having children and devotes herself to helping those she already has, she may have more total offspring who grow up to reproduce themselves"⁷. This hypothesis interprets early reproductive cessation as an adaptive response to prolonged infant dependency. In support, there are some wild species in which females often live well beyond their last pregnancy, including Japanese macaques¹⁰, chimpanzees¹⁰, elephants¹¹, and pilot and killer whales^{2,12}. In all of these species, offspring require extended maternal care.

Good mother theorists have debated which targets of nepotism favour menopause. Is it the last-born offspring⁵, grown daughters and their children^{6,8,13}, or an individual's entire genetic clan⁹? Much attention has been paid to grown daughters and their young, with quantitative models of traditional societies both disputing^{14,15} and affirming¹³ that the fitness benefits of helping to rear grandchildren can exceed the costs of reproductive cessation. Now, Packer *et al.*¹ introduce a comparative perspective to the good mother debate.

The Gombe baboons and Serengeti lions have been studied continuously for over 30 years, beginning with Jane Goodall and George Schaller, respectively. Two to ten times each month, a census is made of recognizable animals, and births and deaths are recorded. The data reveal that female baboons live up to 27 years and female lions up to 17 years. However, fertility drops precipitously when baboons are about 20 years old and the lionesses around 13. Because juvenile baboons require at least two years of maternal care and lion cubs at least one, in both species the mothers can live long enough after ceasing reproduction to rear their last-born young. Thus, lion-cub survival does not decline with maternal age, and, in baboons, survival of infants from old mothers (19–24 years) is just slightly less than those from younger mothers.

Packer *et al.* found that in neither species do post-reproductive females consistently improve the reproductive performance of grown daughters. Whether a female baboon's mother was alive did not affect her age at puberty, interbirth interval, rate of successful pregnancy or survival of her infants to their first birthday. In lions, a

female's litter size and first-year survival of her cubs was also the same regardless of whether her mother was dead, or alive and post-reproductive. Cub survival was improved by the presence of a reproducing grandmother, however, due to better nourishment (females nurse their grandcubs).

The authors interpret their data as support for the senescence theory. They argue that those few lions and baboons that live to old age make a negligible contribution to fitness. Therefore, selection cannot stop their reproductive machinery from deteriorating. Packer *et al.* believe that senescence also accounts for the timing of menopause — by extrapolating from the observed relationship between mortality and maternity in baboons and lions, and assuming a childhood dependency of 10 years in humans, they calculate that the expected lifespan of women who have reached 40 years (when reproduction starts to decline) would be 58–65 years.

Packer *et al.* argue that reproductive cessation in female mammals is a non-adaptive by-product of life-history patterns, and herein lies my only disagreement. Senescence is something that natural selection cannot prevent. Although senescence is non-adaptive, menopause is not. Senescence is inevitable, so in species with prolonged infant dependency, females have evolved a precisely choreographed, adaptive counter-strategy. They cease ovulating abruptly, when they are still young enough to assist kin that would not survive or successfully reproduce without their mother^{5–7,9}. Males are usually not essential to the survival of offspring, so their fertility deteriorates

gradually, solely due to senescence.

Female lions and olive baboons survive only a few years after ceasing to reproduce — apparently, their main targets of maternal nepotism are last-born offspring. But women routinely live to nearly twice the average age at which menopause occurs, implying that children are dependent for a greater proportion of their lives (naïve juveniles in the other species (that is, longer than 10 years). This also suggests that a broad array of relatives benefit from the accumulated wisdom, status and resources of post-reproductive women. □

Paul W. Sherman is in the Section of Neurobiology and Behavior, Cornell University, Ithaca, New York 14853, USA.

1. Packer, C., Tatar, M. & Collins, A. *Nature* **392**, 807–811 (1998).
2. Austad, S. N. *Why We Age* (Wiley, New York, 1997).
3. vom Saal, F. S. & Finch, C. E. in *The Physiology of Reproduction* Vol. 2 (eds Knobil, E. *et al.*) 2351–2413 (Raven, New York, 1988).
4. Symons, D. *The Evolution of Human Sexuality* (Oxford Univ. Press, 1979).
5. Williams, G. C. *Evolution* **11**, 398–411 (1957).
6. Dawkins, R. *The Selfish Gene* (Oxford Univ. Press, 1976).
7. Nesse, R. M. & Williams, G. C. *Why We Get Sick* (Random House, New York, 1994).
8. Lancaster, J. B. & King, B. J. in *In Her Prime* 2nd edn (eds Kerns, V. & Brown, J. K.) 7–15 (Univ. Illinois Press, Urbana, IL, 1992).
9. Alexander, R. D. *Univ. Michigan Mus. Zool. Spec. Publ.* **1**, 1–38 (1990).
10. Takahata, Y., Koyama, N. & Suzuki, S. *Primates* **36**, 169–180 (1995).
11. Laws, R. M., Parker, I. S. C. & Johnstone, R. C. B. *Elephants and Their Habitats* (Oxford Univ. Press, 1975).
12. Marsh, H. & Kasuya, T. *Rep. Int. Whaling Comm. Spec. Issue* **8**, 57–74 (1986).
13. Hawkes, K., O'Connell, J. F., Blurton Jones, N. G., Alvarez, H. & Charnov, E. L. *Proc. Natl Acad. Sci. USA* **95**, 1336–1339 (1998).
14. Hill, K. & Hurtado, A. M. *Hum. Nature* **2**, 313–350 (1991).
15. Rogers, A. *Evol. Ecol.* **7**, 406–420 (1993).

Protein folding

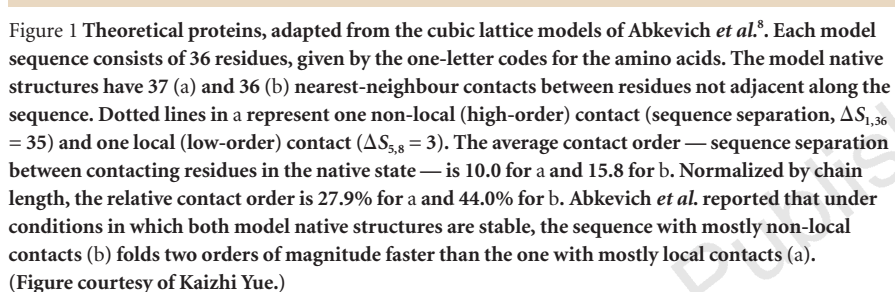
Matching speed and locality

Hue Sun Chan

How does a protein fold? What is the folding code, and how do interactions between amino-acid residues determine the native structure of a protein? Some theoreticians have attempted to tackle these complex problems of molecular recognition and self-assembly by simulations using simple lattice models (Fig. 1, overleaf) — proteins are modelled as chains configured on regular lattices, and, because they lack high-resolution atomic representations, these are models for 'generic proteins'. They try to rationalize general trends in real proteins, but they do not address properties specific to, say, lysozyme or cytochrome *c*. So, theoreticians and experimentalists in the folding field are dealing with different objects — generic versus specific proteins. How can the two be reconciled?

The experimentally observed folding times for different proteins can differ by

more than nine orders of magnitude—from tens of microseconds to hours. Can these be rationalized by models for generic proteins? Unfortunately, theoreticians have not yet reached a consensus on the main determining factor that gives rise to this tremendous diversity in folding rates. In this month's *Journal of Molecular Biology*, Plaxco, Simons and Baker¹ provide much-needed systematic comparisons between theory and experiment. By statistical analysis of the correlations between measured folding rates and the various determinants proposed by theoreticians, they have identified a main structural factor that governs folding speed for a set of non-homologous, single-domain proteins that fold with essentially two-state kinetics. Remarkably, they find this to be the average sequence separation between contacting residues in the native state, normalized by the length of the protein chain — an

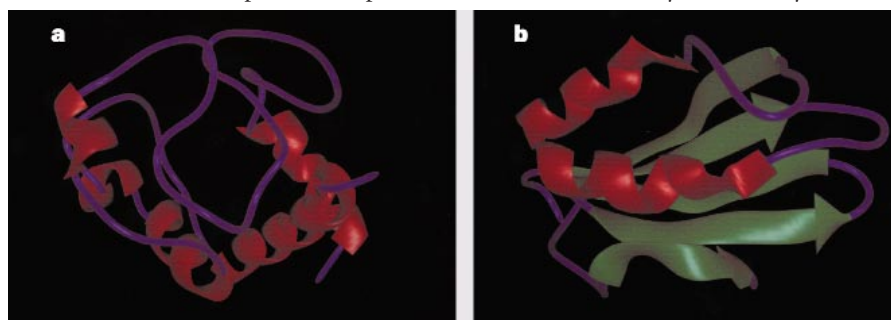


Plaxco *et al.* use their own folding kinetics data, as well as results from other groups. Figure 2 shows the native structures of two proteins in their data set, cytochrome *c* and acyl phosphatase. Although they are about the same size, the folding rates of these two proteins under similar external conditions differ by more than four orders of magnitude. Plaxco *et al.* suggest that the key to this difference in rate is chain topology — the relatively simple topology of cytochrome *c* (with many local contacts) can be formed much more rapidly than the complex topology of acyl phosphatase, which involves the many non-local contacts that make up its β -sheet. For a data set of 12 proteins (which were selected from a larger set to exclude all traces of statistical correlation produced by sequence similarity), Plaxco *et al.* find a significant correlation between relative contact order and folding speed. Also interesting is that native stability and chain length, which have been thought to be important for folding kinetics, fail to show significant correlations with folding rates.

Many theoreticians believe in a step-by-step approach to solving the protein-folding problem. They begin by trying to understand the mechanisms by which protein thermodynamics and folding rates are related to low-resolution properties such as overall chain topology, percentage of hydrophobic residues, strength of intrachain interactions, amount of helices and sheets, and so on. These studies are expected to generate increasingly refined questions and experiments. Indeed, valuable insights have been gained by these approaches, especially with regard to how protein-like behaviours can arise from general properties of chain molecules.



Theoreticians have much to learn from the work of Plaxco *et al.*¹, because their statistical analysis comes close to meeting the generic protein on theoreticians' ground. On the face of it, the new study vindicates theories which predict that local interactions increase the speed of folding^{5,7}, and casts doubt on theories predicting the opposite^{8,9} (Fig. 1). But on closer examination the relation between different theories, and the correspondence between theory and experiment, are more complex. For example, owing to different modelling assumptions, different conclusions have been reached by studies based on the same lattice model^{16,7}. In another instance, two theories^{5,8} agree that non-local interactions increase the thermodynamic stability of the native structure, but come to opposite conclusions about whether local interactions speed up or slow down folding.



762

nant of folding rate — a far less significant factor than contact order. Furthermore, the expected correlation between stability and relative contact order is lacking, leading one to wonder whether any of the theories is directly comparable to the data set of Plaxco *et al.*

The problem seems to be that most of the model sequences studied by theoreticians have a higher tendency to get stuck — kinetically trapped in non-native conformations during folding^{10–12} — than the real, single-domain proteins of Plaxco *et al.* For many model sequences, the key to rapid folding is to overcome kinetic traps. Higher native stability would allow folding to occur at a higher simulation temperature, making it easier to escape from traps. This is the basis for the predictions^{8,9} that non-local interactions cause fast folding, because non-local contacts tend to increase native stability in these models and higher temperatures can be used to simulate folding of more stable sequences with more non-local contacts (Fig. 1)⁸. However, the proteins studied by Plaxco *et al.* do not have much problem with kinetic traps¹², so they should be described by models that minimize the effects of traps. Future modelling will need to capture this generic feature of single-domain, two-state proteins.

It goes without saying that the new statistical analysis does not account for all of the factors that may contribute to folding speed. For example, mutated proteins sharing a common topology (and, thus, the same relative contact order) can fold at different rates: in one case¹³, the variations cover a range of about one order of magnitude. However, if the general picture suggested by the analysis is correct, these other effects should also support the idea that stronger local interactions produce faster folding. This should be testable by further experiments. In the meantime, the work of Plaxco *et al.* is one of many signs that, in the community of protein folders, an increasing constructive feedback between theory and experiment is under way. □

Hue Sun Chan is in the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-1204, USA.
e-mail: chan@maxwell.ucsf.edu

- Plaxco, K. W., Simons, K. T. & Baker, D. *J. Mol. Biol.* **277**, 985–994 (1998).
- Viguera, A.-R., Villegas, V., Aviles, F. X. & Serrano, L. *Folding and Design* **2**, 23–33 (1997).
- Muñoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. *Nature* **390**, 196–199 (1997).
- Viguera, A.-R. & Serrano, L. *Nature Struct. Biol.* **4**, 939–946 (1997).
- Gō, N. & Taketomi, H. *Proc. Natl Acad. Sci. USA* **75**, 559–563 (1978).
- Šali, A., Shakhnovich, E. & Karplus, M. *Nature* **369**, 248–251 (1994).
- Unger, R. & Moulton, J. *J. Mol. Biol.* **259**, 988–994 (1996).
- Abkevich, V. I., Gutin, A. M. & Shakhnovich, E. I. *J. Mol. Biol.* **252**, 460–471 (1995).
- Govindarajan, S. & Goldstein, R. A. *Proteins Struct. Funct. Genet.* **22**, 413–418 (1995).
- Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. *Proteins Struct. Funct. Genet.* **21**, 167–195 (1995).
- Klimov, D. K. & Thirumalai, D. *Phys. Rev. Lett.* **76**, 4070–4073 (1996).
- Chan, H. S. & Dill, K. A. *Proteins Struct. Funct. Genet.* **30**, 2–33 (1998).
- Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L. & Oas, T. G. *Nature Struct. Biol.* **4**, 305–310 (1997).
- Mines, G. A., Pascher, T., Lee, S. C., Winkler, J. R. & Gray, H. B. *Chem. Biol.* **3**, 491–497 (1996).

Spongiform encephalopathies

The prion's perplexing persistence

Adriano Aguzzi and Charles Weissmann

Since prion diseases were identified, we have come to consider such infections as an irrevocable death sentence preceded by severe clinical disease. Take, for example, the cases of iatrogenic Creutzfeldt–Jakob disease elicited by minute traces of the infectious agent, administered through contaminated surgical instruments, hormone preparations or tissue transplants¹. On page 770 of this issue, however, Race and Chesebro² report that infectivity can persist long-term in brain and spleen tissue, without causing the development of clinical symptoms.

There were reasons to believe that exposure to infectious prions may not necessarily lead to clinical disease. For example, mice lacking *Prnp* (the gene that encodes the normal prion protein, PrP^C) are completely immune to clinical disease. They cannot replicate prions, and the traces of infectivity that were occasionally found in the brain several months after injection^{3,4} have been attributed to persistence in the absence of replication⁴. Moreover, mice that contain half the normal level of PrP^C take a very long time to develop symptoms after intra-

cerebral inoculation⁵. Finally, if immunodeficient mice are inoculated peripherally with the scrapie agent, they do not fall ill, despite the occasional presence of large prion loads in their brains⁶.

Another example in which exposure to infectivity does not always lead to disease is the epidemiology of bovine spongiform encephalopathy (BSE). Typically, only single animals from affected farms develop disease, even though the genetic make-up of herds is probably similar, and each cow is likely to have been exposed to similar amounts of prion infectivity. Moreover, although the BSE agent is quite likely the origin of 'new variant' Creutzfeldt–Jakob disease^{7,8}, and although many people in the United Kingdom and Europe may have ingested the infectious agent, only a small minority has developed overt disease to date. Why?

One reason could be that the infectious agent fails to penetrate the organism — perhaps because predisposing factors, such as lesions in the digestive tract, are required. Alternatively, prions may not transfer from the site of entry to the central nervous system. Or, they may not replicate to the extent

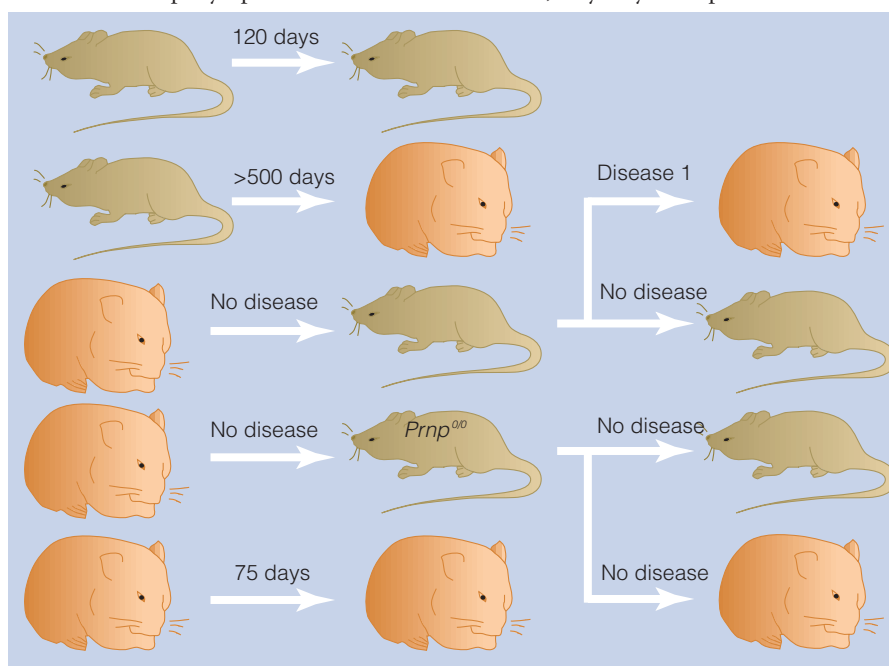


Figure 1 Due to the species barrier between mice and hamsters, prions originating from one species produce spongiform encephalopathy rather quickly when transferred to a further animal of the same species, yet only produce disease — if any — with much longer latency when passed to the other species. This barrier can be abolished by expressing appropriate transgenic prion proteins in the recipient animals^{12,13}.

required for pathogenesis, giving rise to chronic, subclinical infections. In that case, there would be a clinically silent reservoir of prions in otherwise healthy hosts. Because we do not yet have sensitive screening assays for prion infectivity, such reservoirs may be hard to identify and even harder to eradicate.

Race and Chesebro² now show that if mice are inoculated intracerebrally with hamster prions, infectivity is found in their brains and spleens for periods approaching their lifespan. Although the amounts of prion protein are low, and they do not seem to increase significantly with time, they are found very reproducibly. Moreover, the authors detected infectivity only in inoculated wild-type mice — not in *Prnp* knockout mice — indicating that mouse PrP^C is required (Fig. 1).

One may feel inclined to dismiss this unexpected observation as a low-level breach of the tight species barrier that exists between mice and hamsters. What is surprising, however, is that the infectivity found in wild-type mice has the property of hamster and not of mouse prions — it does not cause scrapie in mice, but when it is injected into hamsters they develop symptoms of disease.

Two interpretations are possible: either the infectivity is due to persistence of the hamster prions from the inoculum, or it is the result of low-level replication of the hamster agent in the mouse. Race and Chesebro espouse the first explanation, presumably because the amount of infectivity recovered from the mice is several orders of magnitude lower than that injected. It would have been useful to determine the levels of hamster prion early after inoculation into the mouse because, a few days after mice are intracerebrally inoculated with mouse prions, infectivity in the brain becomes virtually undetectable. Levels can only be measured weeks or months⁹ afterwards, depending on the strains of prion and mouse. If such a decrease in levels of the infectious agent occurs in the experiment of Race and Chesebro, this might indicate that the prions are replicating rather than merely persisting.

Although not discussed by the authors, the possibility of low-level replication of the infectious agent cannot be excluded. If offset by concomitant degradation, this would lead to a low, steady-state level of infectivity. Such a process could be readily explained by a virus or virino hypothesis. But it can also be explained within the framework of the protein-only hypothesis. One would have to imagine that the pathological PrP molecules that constitute the hamster prion (and have a different amino-acid sequence from that of mouse PrP^C) can impart a new conformation to the mouse PrP^C, giving it the properties of a hamster rather than a mouse prion — in other words, primacy of conformation over sequence. In the case of BSE, if the infec-

tious agent responsible is given to various hosts it still gives rise to pathological prion protein (PrP^{Sc}) retaining at least some of its conformation-dependent properties^{4,8}. So, it is important to clarify whether hamster prions are replicating silently in wild-type mice, or whether they are merely persisting.

Race and Chesebro note that it may not only be mice and hamsters that behave in this way when exposed to prions from other species. This could have considerable implications — for example, farm animals that do not contract overt disease after consuming ruminant-derived meat and bone meal may, perhaps, develop a subclinical carrier state. Pigs and chickens that have been fed with cattle-derived bone and meat meal are thought to be safe to eat with respect to BSE, because these animals do not develop disease after oral exposure to bovine prions. But, to the best of our knowledge, bovine prions from BSE-exposed pigs and poultry have never been assayed using calves as 'indicator' animals.

Why, in the experiments of Race and Chesebro, does hamster infectivity not persist in mice that lack the gene encoding PrP^C? It is possible that the immune system is involved. Mice lacking PrP develop a cellular immune response to PrP^C (ref. 10), and, in fact, some of the best monoclonal antibodies to PrP^{Sc} have been derived from such mice¹¹. Perhaps hamster prion infectivity cannot persist in PrP-deficient mice because anti-prion or anti-PrP antibodies wipe it out. In the wild-type mice, however, it could persist owing to their relative immune tolerance to PrP^{Sc}. This hypothesis could be tested by repeating the experiment reported here with mice that do not express wild-type mouse PrP^C and are immunocompromised. Given the current interest in the collusion between prions and immune cells, such experiments are likely to be undertaken sooner rather than later. □

Adriano Aguzzi is at the Institute of Neuropathology, University Hospital of Zürich, CH-8091 Zürich, Switzerland.

e-mail: adriano@pathol.unizh.ch

Charles Weissmann is at the Institute of Molecular Biology, University of Zürich, Höggerberg, 8093 Zürich, Switzerland.

e-mail: weissma@molbio1.unizh.ch

1. Brown, P., Preece, M. A. & Will, R. G. *Lancet* **340**, 24–27 (1992).
2. Race, R. & Chesebro, B. *Nature* **392**, 770 (1998).
3. Sakaguchi, S. *et al. J. Virol.* **69**, 7586–7592 (1995).
4. Sailer, A., Büeler, H., Fischer, M., Aguzzi, A. & Weissmann, C. *Cell* **77**, 967–968 (1994).
5. Büeler, H. *et al. Mol. Med.* **1**, 19–30 (1994).
6. Klein, M. A. *et al. Nature* **390**, 687–690 (1997).
7. Hill, A. F. *et al. Nature* **389**, 448–450 (1997).
8. Bruce, M. E. *et al. Nature* **389**, 498–501 (1997).
9. Kimberlin, R. H. & Walker, C. A. *J. Gen. Virol.* **42**, 107–117 (1979).
10. Brandner, S. *et al. Proc. Natl Acad. Sci. USA* **93**, 13148–13151 (1996).
11. Korth, C. *et al. Nature* **390**, 74–77 (1997).
12. Scott, M. *et al. Cell* **59**, 847–857 (1989).
13. Prusiner, S. B. *et al. Cell* **63**, 673–686 (1990).

Daedalus

Penetrating gaze

There's an audio-analyser for telephone conversations, which is alleged to reveal when the speaker is lying. Duplicity, it seems, causes detectable changes in certain voice-frequencies. But in normal life, even verbal communication is largely visual. It is claimed that 55% of a speaker's impact derives from his appearance and body language, 38% from his tone of voice, and only 7% from what he actually says. So Daedalus is taking the obvious next step. He is devising an image-analyser to tell if a speaker on video is lying.

DREADCO technicians have set up a special studio for the purpose, using high-resolution cameras with an enhanced frame-rate. Volunteers, from clergymen through accountants to car salesmen and juvenile criminals, are being invited to converse on a wide range of topics, from philosophical speculation to embarrassing personal confessions, lying whenever they feel they can get away with it. A panel of detectives, psychiatrists and tax inspectors is studying the footage, and the DREADCO team are correlating their conclusions with details from the tape.

Each frame is analysed by modern pattern-recognition software, to identify the face and body of each speaker; these are then being searched for specific clues. Daedalus expects a liar to show subtle inconsistencies between face and body; blink-rate and gaze direction may conflict with hand-movements or shifting stance, and both may conflict with verbal stresses in the audio channel. A full spatial Fourier analysis of all movements should reveal the key stigmata of dishonesty, and with luck will be able to place it on a spectrum from trivial detail-bending to major fraud. The most reliable audiovisual indicators of dishonesty will then be simplified, enhanced, and adapted for normal TV and closed-circuit video signals.

DREADCO's 'Lying Eye' program will sell like hot cakes, starting in the rapidly growing video-conferencing market. Those base suspicions of tele-workers, long-distance contacts, and all the other evasive types who won't attend proper meetings, will be rapidly tested. The next obvious market will be the law — both for police interviews (now routinely videotaped to avoid claims of trickery) and in the courts themselves. Many doubtful cases will collapse spontaneously when it turns out that the plaintiff, the defendant, the witnesses on both sides, and the defending and prosecuting lawyers themselves, are all lying.

David Jones

Merz's maths

Mario Merz seeks to take art beyond the aesthetic and into the realms of experimentation. A favourite subject of his has been the mathematics of growing populations.

Martin Kemp

Mario Merz, born in Italy in 1925, has been one of the pioneering artists who have aspired to transform the art gallery into something resembling a laboratory for the experimental interaction of visual imagination and intellectual enquiry. A major focus of his own intellectual enquiry has been the Fibonacci series of numbers: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55....

Whether working with paint on canvas or with the materials he uses in his constructions — stone, clay, wood, fruit, metal, liquids, glass, neon tubes, fabric, paper — Merz presents a consistency of purpose that can be recognized in his dialogues between shape, structural principles, physical forces, the enclosure and expansion of space and the mathematics of growth, in the context of human existence.

Put in this way, Merz's agenda seems both pretentious and unrealizable. But a series of themes has opened windows on to how it might be realized by a programme extending beyond his own production. Two such themes are his "tables" and his "igloos".

His tables, closer to Japanese than western examples, envisage a population growing in accordance with Leonardo da Pisa's famous thirteenth-century series, which was formulated as a model of rabbit populations.

It is as if we are planning a canteen for proliferating bands of factory workers. As Merz wrote in his accompanying poem:



Merz's *Tavole con le Zampe Diventano Tavoli* (Boards with Legs Become Tables), 1974.

"I reject the linear, one by one, or assembly-line fabrication of spaces...

The space grows like a bunch of grapes.

For a growing number of people it is necessary to make tables that grow like a bunch of grapes.

For one person.

For two people then..."

He takes the series as far as 34, but its potential is boundless.

The Fibonacci series signifies life beyond the apparent inertness of numbers. The lower numbers remain implicated within the larger terms, and the progression embodies

the principles of organic growth: "the numerals in PROLIFERATION have in themselves the power of SUCCESSION or PROPAGATION".

Expressed in such figures as a cone or logarithmic spiral, the series exhibits a power of propagation to infinity through morphologies that are self-similar at every stage.

The structure of Merz's igloos — the form of primitive houses in many cultures — expresses the interplay between growth and structure in terms of volume.

The number of wall units needed to fill the interstices — whether Merz's stone, clay and glass or the snow and natural materials of actual dwellings — both enact a process of numerical growth in each instance, and register the exponential increase with each successive enlarging of the radius of the structures.

The igloos strike a balance between gathered strength at their summits and precarious equilibrium as they reach out to cover ever larger areas. The tension is nicely encapsulated in the paradoxical aphorism of the Vietnamese leader, General Giap, who inspired one of Merz's igloos: "If the enemy concentrates it loses ground; if the enemy disperses it loses force." Such paradoxes, for Merz, permeate not just structure and growth but also the operations of human society. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.



Merz's *Igloo di Stoffa*, 1968–81.

Creating corners in kitchen sinks

When a vertical jet of liquid from a nozzle hits a flat surface, as in tap water striking the kitchen sink, a discontinuity appears in a ring created by the flow. At this deformation the depth of the water alters abruptly (the 'circular hydraulic jump'¹⁻⁵) at some distance from the jet. We have now discovered that if the jet contains a liquid more viscous than water, stationary polygonal patterns form, breaking axial symmetry. The sharp corners of the polygons carry a large radial flux, while the sides generate resistance to the stream.

To account for the appearance of these remarkable structures in a simple flow, we followed a classic idea¹ to design the experiment shown in Fig. 1a. Close to the jet, the fluid layer is thin (of height h_{int}) and the motion is rapid, whereas further away the layer height is increased by an order of magnitude to h_{ext} and the fluid moves correspondingly more slowly.

The transition between these two types of motion occurs within a surprisingly short distance — a millimeter or so. The horizontal surface was a large disc (diameter 36 cm) made of glass so that the jump can be observed from below. The fluid going over the rim around the disc is collected and recirculated, and the rim can be raised or lowered to control h_{ext} (ref. 4).

The circular jump can occur in two states that have different cross-sectional flow patterns. When h_{ext} is small, we observe a type I pattern, in which the surface velocity is outward everywhere and a separation bubble is present at the bottom just outside the jump. On increasing h_{ext} , the jump becomes steeper, until, at a critical value of h_{ext} , fluid outside the jump flows inwards and breaks over the edge of the jump⁴.

When a new steady state (type II pattern) is reached, a surface eddy has formed. This eddy, called a 'roller' or a 'surfing wave', has the geometry of a floating torus surrounding the jump (Fig. 1a). The outgoing flow passes under the roller. This maintains the rotation of the roller and an inward surface velocity near the jump.

In experiments with ethylene glycol (antifreeze), which is about ten times more viscous than water, the circular type II state (Fig. 1b) frequently undergoes spontaneous breaking of its azimuthal symmetry into a stationary polygonal shape (Fig. 1c). Rather than displaying the weak angular deformations generally seen in fluids, the jumps form clear corners and edges that are often straight (Fig. 1c,d) but which can curve outwards or inwards.

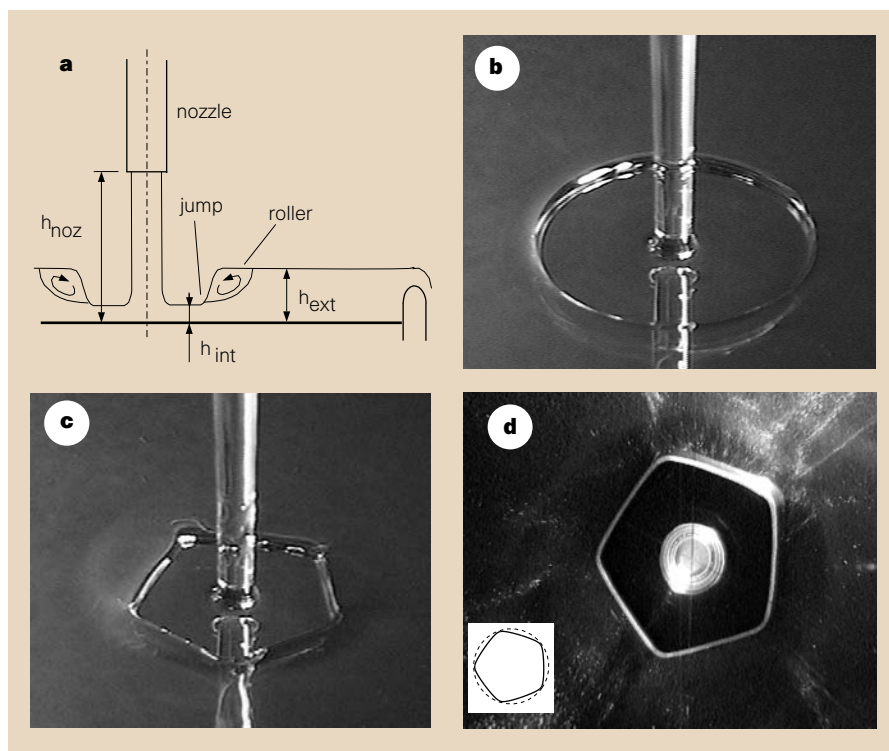


Figure 1 The hydraulic jump experiment. A cross-sectional flow structure (type II pattern) is shown in **a**; **b-d** are video images of three stationary states. The fluid is ethylene glycol (antifreeze; about 99% pure) and typical flow rates, Q , are 30–50 ml s⁻¹. The external height, h_{ext} , is increased from **b** to **c**, causing the familiar circular jump in **b** to change spontaneously to polygons with many corners (not shown), then to a pentagon (**c**), viewed from underneath in **d**. In **d**, the bright circle in the centre is the vertical jet; the corners of the surrounding polygonal jump are clearly evident. The inset shows one solution to our model.

These states are robust — they reappear if the flow is temporarily disrupted or scrambled. However, several polygon states can be stable for the same h_{ext} .

There is an interval of h_{ext} in which, for example, a pentagon can be deformed into a hexagon through provoking another corner by locally dragging a side of the jump with a pin. Corners prefer to be equally spaced: if not, a drift and an adjustment in the azimuthal direction occurs over a long period (sometimes lasting for hours), even though the radial flow is rapid.

As h_{ext} is raised from the type I configuration, the wave breaking may create many corners: polygons with up to 14 corners are created by changing the flow rate Q , the height of the nozzle outlet h_{noz} , and the kinematic viscosity ν . As h_{ext} is increased further, the corners disappear, usually one by one, and the circumference of the jump becomes smaller until the jump eventually 'closes'.

There is a considerable hysteresis effect in the sense that several polygons can be stable at the same flow parameters. Interestingly, all these coexisting polygons seem to have exactly the same corner shape.

Inside the roller, a slow spiralling flow develops towards the corner, where the roller breaks up.

How can such unusual structures occur in these simple flows? We have derived a model which assumes that radial forces on the roller, namely the hydrostatic inward pressure and the outward shear force from the radial flux, balance at any angle, even when the radial flux is allowed to vary. This balance results in a relation between the width of the roller and the radial flux at a given angle.

If we further assume that the roller possesses a weak line tension, and therefore that the shape chosen is the one with minimal circumference for a given imposed flux, we get a simple variational problem that can be solved numerically (one of the computed shapes is shown in Fig. 1d). The model can be treated as non-dimensional so that the only parameter is the dimensionless flux ϕ , which equals $\nu Q / (g(h_{\text{ext}}^2 - h_{\text{int}}^2))$.

The circular jump exists for all values of ϕ , and more corners arrive as it increases. There is overlap of the intervals of ϕ for which various polygons exist, so that several polygons can be found for the same system

parameters, just as we find in the experiments described here.

Clive Ellegaard, Adam Espe Hansen, Anders Haaning, Kim Hansen, Anders Marcussen, Tomas Bohr, Jonas Lundbek Hansen, Shinya Watanabe
Center for Chaos and Turbulence Studies,
The Niels Bohr Institute, Blegdamsvej 17,
DK-2100 Copenhagen, Denmark

1. Rayleigh, Lord *Proc.Roy.Soc.Lond. A* **90**, 324–328 (1914).
2. Craik, A., Latham, R., Fawkes, R. & Gribbon, P. *J. Fluid Mech.* **112**, 347–362 (1981).
3. Bohr, T., Dimon, P. & Putkaradze, V. *J. Fluid Mech.* **254**, 635–648 (1993).
4. Bohr, T., Ellegaard, C., Hansen, A. E. & Haaning, A. *Physica B* **228**, 1–10 (1996).
5. Bohr, T., Putkaradze, V. & Watanabe, S. *Phys. Rev. Lett.* **79**, 1038–1041 (1997).

Electron microscopy image enhanced

One of the biggest obstacles in improving the resolution of the electron microscope has always been the blurring of the image caused by lens aberrations. Here we report a solution to this problem for a medium-voltage electron microscope which gives a stunning enhancement of image quality.

Even today, more than 60 years after the invention of the transmission electron microscope¹, the point resolution is still limited by the spherical aberration of its objective lens. Up until now, the only way to improve matters was to impose a numerical correction for this spherical aberration that was based either on a series of images of the object taken under variable objective-focus conditions² or on the application of holographic techniques³.

Fifty years ago, Scherzer⁴ suggested that the two principal axial aberrations, chromatic and spherical, could be corrected by electrostatic or magnetic multipole elements, but this proved to be beyond the technology available at that time. Our solution for spherical aberration correction derives from a much more recent suggestion by Rose⁵, and depends on two electromagnetic hexapoles and four additional lenses.

The principle by which correction is achieved is based on the fact that the primary aberrations of second order from the first hexapole are compensated by the second hexapole element. However, the two hexapoles additionally induce a residual secondary third-order spherical aberration which is rotationally symmetric⁶ and proportional to the square of the hexapole strength.

The appertaining coefficient of spherical aberration is of the opposite sign to that of the objective lens. In this way, the spherical aberration of the entire system can be

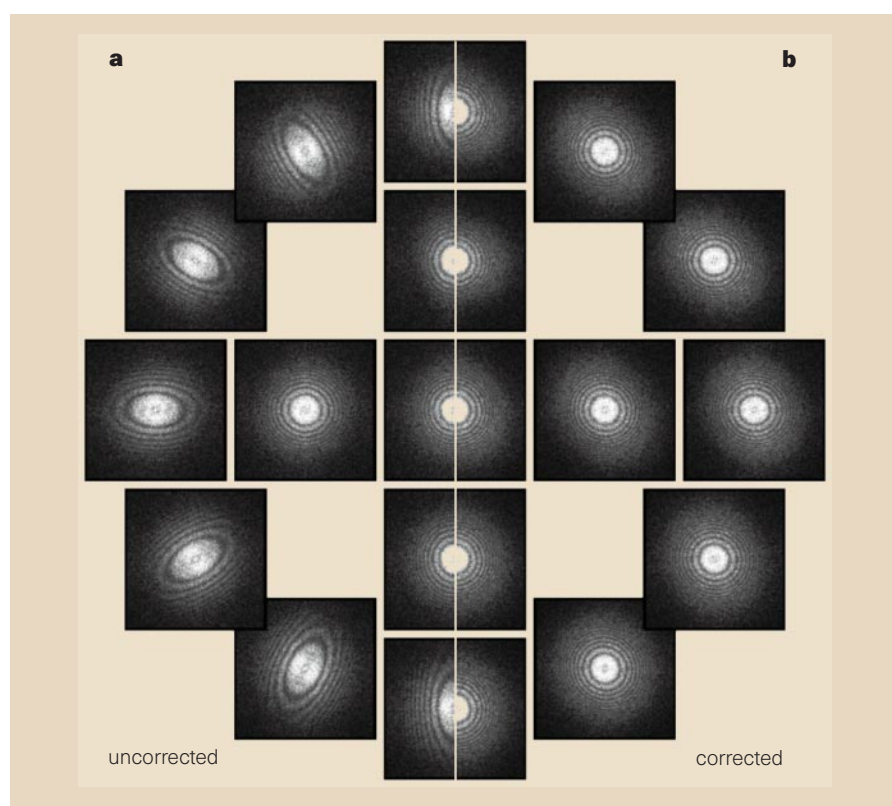


Figure 1 Diffractogram tableau of the microscope. **a**, Uncorrected, and **b**, after correction for spherical aberration and proper alignment. The beam tilt angle is 10.8 mrad in both cases and the azimuthal angles vary between 0 and 2π in steps of $\pi/6$. The essentially identical shape of the diffractograms in **b** indicates the vanishing spherical aberration.

compensated by exciting the hexapoles appropriately⁷.

In the transmission electron microscope, off-axis aberrations are as important as axial ones. Therefore, in order to maintain a finite field of view, a semi-aplanatic objective lens system had to be constructed which fulfils Abbe's sine condition. This means that it is not only free of spherical aberration but also of off-axis coma and parasitic axial aberrations resulting from misalignment of the optical system.

Spherical aberration correction is not enough to improve the resolution of the microscope. It is necessary to compensate as well for the parasitic second-order axial aberrations, coma and threefold astigmatism, and adequately to suppress the non-

spherical axial third-order aberrations, star aberration and fourfold astigmatism. For this purpose, the aberration coefficients are determined by means of an extended version of the diffractogram tableau method⁸.

With the corrector switched off, a diffractogram tableau of an amorphous sample is recorded digitally (Fig. 1a) and evaluated on-line in terms of induced defocus and twofold astigmatism. Switching on the corrector introduces misalignment aberrations and an appropriate alignment procedure has to be carried out; this semi-automatic alignment takes about 15 minutes.

The diffractogram tableau of the corrected and aligned system is shown in Fig. 1b. All diffractograms are of roughly the same shape, indicating the properties

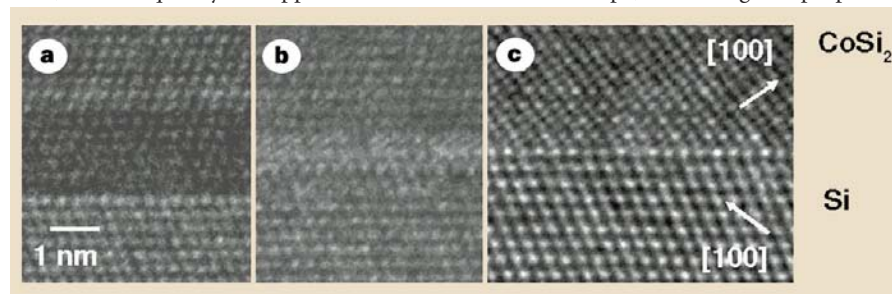


Figure 2 Structure images of an epitaxial Si(111) CoSi₂ interface demonstrating the production of image artefacts by the effect of contrast delocalization due to spherical aberration. Images in **a** and **b** were taken in the uncorrected microscope at Scherzer defocus and Lichte defocus, respectively. **c**, Image taken in the aberration-corrected state at Scherzer defocus does not show any delocalization ($C_s = 0.05$ mm).

of an aplanatic objective lens. Even for beam tilt angles as high as 30 mrad, the values of defocus and two-fold astigmatism induced by the residual axial aberrations remain small.

The basic instrument used for this development was a standard 200 kV Philips CM 200 ST equipped with a field emission gun. The corrector increases the column height by 24 cm but leaves the other operating modes and functions of the microscope unaffected. With this modified microscope, we were able to demonstrate the correction of the spherical aberration and to realize an improvement of the point resolution from 0.24 nm to better than 0.14 nm.

The main advantage of spherical aberration correction is that structure-imaging artefacts due to contrast delocalization can to a great extent be avoided. These artefacts have turned out to be a major obstacle for the application of instruments with field emission guns in defect and interface studies⁹. Contrast delocalization arises from the width of the aberration discs belonging to the individual diffracted electron waves whose diameter increases with the spherical aberration.

Figure 2a shows, in a cross-sectional preparation, the interface of CoSi₂ grown epitaxially on Si(111) seen along the <110> direction. The image was taken under Scherzer defocus without correction. At the interface, an approximately 2-nm-broad region of darker contrast can be seen. The width depends on the defocus value reaching a minimum at the Lichte defocus¹⁰ (Fig. 2b). If the structure is imaged in the aberration-corrected condition (Fig. 2c), the delocalization has essentially disappeared and the interface is atomically sharp.

Maximilian Haider*, Stephan Uhlemann*, Eugen Schwan

European Molecular Biology Laboratory,
Postfach 102209, 69012 Heidelberg, Germany

*Present address: CEOS GmbH, Im Neuenheimer
Feld 519, 69120 Heidelberg, Germany

Harald Rose

Institut für Angewandte Physik, Technische
Hochschule Darmstadt,
64289 Darmstadt, Germany

Bernd Kabius, Knut Urban

Institut für Festkörperforschung,
Forschungszentrum Jülich GmbH,
52425 Jülich, Germany

1. Knoll, M. & Ruska, E. *Z. Physik* **78**, 318–339 (1932).

2. Kirkland, E.J. *Ultramicroscopy* **15**, 151–172 (1984).

3. Lichte, H. *Ultramicroscopy* **20**, 293–304 (1986).

4. Scherzer, O. *Optik* **2**, 114–132 (1947).

5. Rose, H. *Optik* **85**, 19–24 (1990).

6. Beck, V.D. *Optik* **53**, 241–255 (1979).

7. Haider, M., Braunshausen, G. & Schwan, E. *Optik* **99**,
167–179 (1995).

8. Weimlin, F., Weiss, K., Schiske, P., Kunath, W. & Herrmann,
K.-H. *Ultramicroscopy* **3**, 49–60 (1977).

9. Thust, A., Coene, W.M.J., Op de Beek, M. & Van Dyck, D.
Ultramicroscopy **64**, 211–230 (1996).

10. Lichte, H. *Ultramicroscopy* **38**, 13–22 (1991).

Point vortices exhibit asymmetric equilibria

The equilibrium patterns formed by interacting vortices have been puzzled over in a variety of contexts since Kelvin's theory of vortex atoms¹ was debunked by quantum mechanics. These patterns have, for example, appeared in atmospheric and oceanographic flows^{2,3} and in rotating superfluid helium^{4,5}. For a particular mathematical model, the point vortex equations, a catalogue of equilibria was drawn up some twenty years ago⁶. We have revisited the point vortex model by a different approach and have discovered a host of new equilibrium states, including asymmetric patterns devoid of rotational or reflectional symmetry.

The dynamical equations considered are

$$\frac{dz_{\alpha}^*}{dt} = \frac{\Gamma}{2\pi i} \sum_{\beta=1}^{N'} \frac{1}{z_{\alpha} - z_{\beta}} \quad (1)$$

where the vortices, all of strength Γ , are parallel lines that intersect a plane of flow, thought of as the complex plane, at points, z_{α} , $\alpha = 1, \dots, N$. The asterisk denotes complex conjugation; the prime indicates $\beta \neq \alpha$. For a configuration rotating uniformly with angular frequency ω , $z_{\alpha}(t) = z_{\alpha}(0)e^{i\omega t}$, and equation (1) gives the algebraic system:

$$z_{\alpha}^* = \sum_{\beta=1}^{N'} \frac{1}{z_{\alpha} - z_{\beta}} \quad (2)$$

We have rescaled such that $2\pi\omega/\Gamma = 1$. Because of the appearance of complex conjugation, even determining the total number of solutions of equation (2) for a given N is not a simple matter.

We now solve equation (2) by a novel numerical approach. Starting from an equilibrium with N vortices, we first determine all co-rotating points, z , by solving

$$z^* = \sum_{\alpha=1}^N \frac{1}{z - z_{\alpha}} \quad (3)$$

which we may call Morton's equation⁷ as he first studied it for N values of 3 and 4.

Given the equilibrium and a co-rotating point, we apply Newton's method to solve the system of $N+1$ equations

$$z_{\alpha}^* = \sum_{\beta=1}^{N'} \frac{1}{z_{\alpha} - z_{\beta}} + \frac{\epsilon}{z_{\alpha} - z} \quad (4)$$

for the instantaneous positions $\alpha = 1, \dots, N$, together with equation (3), as ϵ varies from 0 to 1. This determines an equilibrium of $N+1$ vortices, with N vortices of strength 1 and one of strength ϵ , with the same angular velocity.

We have no guarantee that solutions exist continuously in ϵ . Numerical results indicate that sometimes, when ϵ is incremented, there are no solutions close to the one at hand. When it succeeds, we refer to this process as 'growing' an equilibrium with $N+1$ vortices from one with N vortices and a co-rotating point. Similarly, we start with an equilibrium of $N+1$ vortices and gradually reduce the strength of one of these from 1 to 0, again solving equations (3) and (4) at every stage with ϵ decremented step by step, until we arrive at an equilibrium of N vortices and a co-rotating point.

We have also placed 'seed' vortices at $k > 1$ co-rotating points and 'grown' a new equilibrium with $N+k$ vortices from one with N , and we have similarly reduced k vortices step by step in strength from 1 to 0, producing an equilibrium with $N-k$ vortices and k co-rotating points. Our method tends to produce 'families' of equilibria, namely similar patterns with different values of N .

By using these procedures, we have discovered a large number of new equilibria, many of which depart significantly from the nested-ring type of configuration recorded in the Los Alamos catalogue⁶. For example, some of our families of configurations are

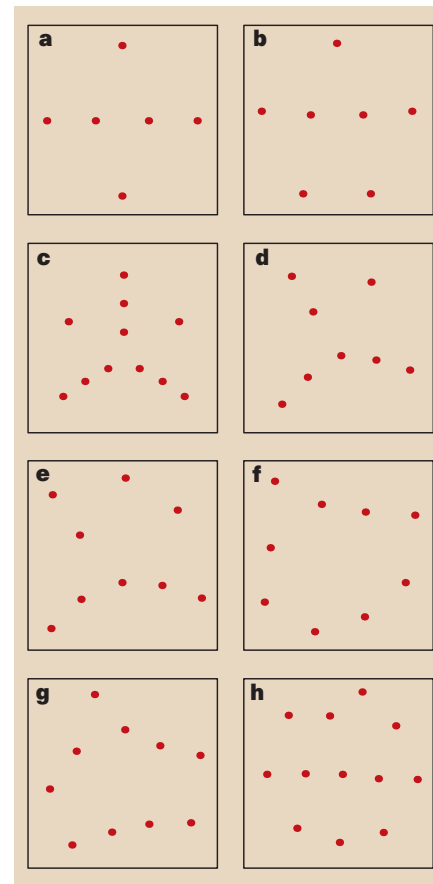


Figure 1 Equilibria of N identical vortices. **a**, Vortices arranged on perpendicular lines, $N = 6$; **b–c**, 'irregular' patterns with a symmetry axis, for $N = 7$ or 11; **d–h**, asymmetric patterns, for N values of 8, 9, 10 or 12.

suggestive of an arrangement along perpendicular lines, rather than in rings (Fig. 1a; it is possible to determine these analytically); we also found many other irregular configurations (such as those shown in Fig. 1b,c). We stress that in each case the configuration is of the same type for several different values of N , although only one value of N has been shown here.

We have also revealed many new vortex patterns of the symmetrical, nested-ring type. All of these have so far turned out to be unstable, but their number and geometric richness has surprised us.

Finally, and most importantly, we have identified equilibria that have no rotational or reflectional symmetry at all. Examples for N values of 8, 9, 10 and 12 are shown in Fig. 1d–h, but we believe that such equilibria exist for all $N \geq 8$ (and we have found them for N values of 8–14). Note that the reflection of any of these patterns yields a new equilibrium.

These asymmetric equilibria appear to belong to several families, and we have established relations among some pairs of equilibria for different N by the method of growing and reducing equilibria that we describe here.

In a non-dissipative system, such as point vortices, unstable equilibria play an important role in determining system evolution: as such states are approached, evolution slows down. The existence of unstable, asymmetric equilibria suggests that such configurations will often be observed when a vortex system is started from general initial conditions.

Hassan Aref, Dmitri L. Vainchtein

Department of Theoretical and Applied Mechanics,
University of Illinois at Urbana-Champaign,
216 Talbot Laboratory, 104 South Wright Street,
Urbana, Illinois 61801, USA

1. Thomson, W. *Nature* **18**, 13–14 (1878).
2. Stewart, H. J. Q. *J. Appl. Math.* **1**, 262–267 (1943).
3. Carnevale, G. F. & Kloosterziel, R. C. *J. Fluid Mech.* **259**, 305–331 (1995).
4. Williams, G. A. & Packard, R. E. *Phys. Rev. Lett.* **33**, 280–283 (1974).
5. Yarmchuk, E. J., Gordon, M. J. V. & Packard, R. E. *Phys. Rev. Lett.* **43**, 214–217 (1979).
6. Campbell, L. J. & Ziff, R. M. *Los Alamos Sci. Lab. Rep.* LA-7384-MS (1978).
7. Morton, W. B. *Proc. R. Irish Acad. A* **41**, 94–101 (1933).

Scrapie infectivity found in resistant species

Experimental and epidemiological evidence indicates that bovine spongiform encephalopathy (BSE) has been transmitted to humans^{1–4}, although the mechanism of this transmission is unknown. Hamsters and chickens are clinically resistant to the transmission of BSE, but we report results that raise concern over the possible long-term

Table 1 Persistence of hamster scrapie agent in normal mice depends on expression of mouse PrP

Mice inoculated with hamster scrapie			Recipient hamsters		Recipient mice
			Brain	Spleen	Brain
Mouse	Mouse PrP genotype	Days after inoculation	No. positive/total	Incubation period	No. positive/total
1	+/+	204	4/4	313 ± 112	NT
2	+/+	310	3/3	178 ± 14	NT
3	+/+	574	7/7	140 ± 4	6/6
4	+/+	574	8/8	162 ± 38	6/6
5	+/+	693	8/8	156 ± 20	6/6
6	+/+	693	8/8	168 ± 42	6/6
7	+/+	782	8/8	128 ± 13	NT
8	0/0	112	0/12	–	NT
9	0/0	112	2/12	294–334	NT
10	0/0	217	0/12	–	1/8
11	0/0	217	4/12	296–352	0/8
12	0/0	330	0/7	–	0/8
13	0/0	330	0/8	–	0/8

PrP-null ($PrP^{0/0}$) mice⁷ (provided by J. Manson, Neuropathogenesis Unit, Edinburgh) were backcrossed four times to C57BL/10. The original donor mice were inoculated intracerebrally with 50 μ l of a 1% suspension of hamster scrapie brain containing 1×10^7 ID₅₀ units. Mice were killed after the indicated number of days after inoculation with hamster scrapie. To test for persistent scrapie infectivity, 50 μ l of a 1% suspension of brain or spleen was inoculated intracerebrally into recipient hamsters or mice. NT, not tested.

persistence of infectivity in such clinically persistent species and which may have implications for the control of BSE.

‘Natural’ transmission of BSE seems to result mainly, if not solely, from feeding animals with meat and bonemeal derived from BSE-infected cattle⁵. BSE is transmissible by inoculation or ingestion to mice, sheep, goats, marmosets, mink, pigs, certain members of the cat family and various exotic ungulates.

Interestingly, though, it is not transmissible to hamsters or chickens⁵. This resistance to BSE is borne out by an absence of clinical brain disease and a lack of the hallmark protease-resistant prion protein.

We injected 10^7 ID₅₀ (where ID₅₀ is the half-maximal infectious dose) units of hamster scrapie agent (strain 263K) into the brains of groups of C57BL/10 mice that either did or did not express the mouse gene that encodes the normal version of the prion protein (PrP). Mice are highly resistant to this hamster scrapie strain⁶ and, as expected, none of our mice developed any clinical symptoms of scrapie.

We found, however, that brain and spleen tissue from the PrP-positive mice obtained between 204 and 782 days after inoculation contained scrapie agent that was capable of infecting hamsters but not mice (Table 1). All of the recipient hamsters were positive, although the long incubation periods indicated that the amount of infectivity present was about 10^2 ID₅₀ per gram of tissue, far lower than is typically found in clinical scrapie in hamsters or mice.

By contrast, tissues of mice without a functional PrP gene ($PrP^{0/0}$ mice) were only very rarely able to transmit disease to recipient hamsters. Thus, expression of a normal mouse PrP gene seemed to enable hamster

scrapie agent to persist in the brains and spleens of normal mice, suggesting that perhaps mouse PrP could be functioning as a receptor for the infectious agent⁴. However, there was no evidence for any replication of the hamster scrapie agent in these persistently infected mice. The infectivity recovered remained specific for hamsters rather than mice (Table 1).

Although we have not tested whether similar results would be obtained after oral ingestion, this unexpected and prolonged survival of a foreign scrapie agent raises the possibility that BSE infectivity might persist in various ‘resistant’ species exposed to BSE-contaminated feeds. Of particular concern would be domestic animals such as poultry which are raised for human consumption.

So far, there is no evidence for the secondary transmission of BSE from such resistant species to more susceptible species. However, the results presented here would strongly favour a decision to stop feeding ruminant-derived products to all animal species. Additional experiments should be carried out to detect possible BSE infectivity in clinically normal BSE-exposed animal species. Titration in cattle may be necessary to achieve the sensitivity needed to detect the levels of BSE anticipated in such situations.

Richard Race, Bruce Chesebro

Laboratory of Persistent Viral Diseases,
National Institute of Allergy and Infectious Diseases,
Hamilton, Montana 59840, USA
e-mail: bchesebro@nih.gov

1. Will, R. G. *et al.* *Lancet* **347**, 921–925 (1996).
2. Bruce, M. E. *et al.* *Nature* **389**, 498–501 (1997).
3. Hill, A. F. *et al.* *Nature* **389**, 448–450 (1997).
4. Chesebro, B. *Science* **279**, 42–43 (1998).
5. Collee, J. G. & Bradley, R. *Lancet* **349**, 636–641 (1997).
6. Kimberlin, R. H. *et al.* *J. Gen. Virol.* **70**, 2017–2025 (1989).
7. Manson, J. C. *Mol. Neurobiol.* **8**, 121–127 (1994).

Designer genes and legal briefs

Remaking Eden: Cloning and Beyond in a Brave New World

by Lee M. Silver

Avon/Weidenfeld and Nicolson: 1998.

Pp. 315. \$25, £20

John Cairns

In the nineteenth century, most of what was going on in the sciences was accessible to the whole of the reading public. George Eliot, it is said, was engrossed in the proofs of Darwin's *Origin of Species* on the day it was published; and doctors in the United States had built their own X-ray machines for the treatment of breast cancer within a few months of reading about Röntgen's discovery of X-rays. During the past century, however, science has become so weighed down with facts that most of us can understand it only through the use of some kind of intermediary.

At the same time, science has become increasingly important. Political changes dominated the nineteenth century; science promises to dominate the twenty-first. Because democracy has proved to be, in the long run, the safest political system, it is very important that everyone, not just an intellectual élite, should have access to reliable guides to the science underlying what is happening to the human condition.

Nowhere is this more important than in the matter of genetic engineering, where opportunities for irreversible mischief seem almost limitless. So it is a great pleasure to report that Lee M. Silver's book about the genetic engineering of humans is very good indeed. He has first-hand knowledge of his subject and writes clearly and skilfully. His book covers the ways in which we are now able, or may soon be able, to decide the genetic constitution of our children. The description is in the form of a series of little family histories, some real, some imaginary. Each serves to introduce some particular technology and is accompanied by an account of the relevant sector of molecular biology or developmental genetics.

Some of the problems have already come to light. For example, what should be done with frozen embryos if both parents are killed in an accident? Is it right for a woman to decide to have a second child so that it can provide the marrow transplant that may save the life of a first child who is dying of leukaemia?

Other problems are just around the corner. Should a woman be allowed to bear a daughter who is a clone of herself (and, I might add, if the woman later dies, should her widower be allowed to marry the clone on the grounds that he is, in effect, remarrying his wife)?

Some of the case histories discussed by



Nice body work: products of engineering as foreseen in the film *Brave New World*.

Silver promise to be as divisive as the issue of abortion. He seems particularly worried that genetic manipulation may eventually be able to offer general benefits, such as increased intelligence and resistance to disease. Because of the expense, these benefits will be available only to the rich, and he fears that, after many generations of manipulation, the human population may find itself divided into two distinct species (he calls them 'GeneRich' and 'Naturals') that cannot interbreed. Of course, the separation into rich and poor has been with us since the beginning of civilization, but has been partly relieved by a steady flow between the two groups owing to the fluctuation in people's fortunes. A division based on artificially enhanced intelligence might be far more destructive. (I understood that the British Labour Party, when it came to power after the Second World War, decided to leave untouched the so-called public schools because it felt that an oligarchy based on wealth was bound to be less entrenched than one created by extra education for the cleverest children.)

The book deserves to be widely read, not least because it gives such a lucid account of the science. But I am less worried than Silver about the genetic engineering of humans. Far more important, I think, is the danger if we come to rely exclusively on highly engineered crops, and the danger posed by new microorganisms that terrorists can now design using equipment as compact as the apparatus of the nineteenth-century physicist that could have been brought in by his butler, on a tray. Surely, if there is a forbidden apple in the new Eden, it is most likely to be

found in the genetics of plants or microbes.

I hope that Silver will now look at these other fruits of genetic engineering. In some ways they are harder subjects because they do not concern simply the morality and legality of what can be done. But they are a more likely way for mankind to get into trouble than just by meddling with the genes of some of the richer members of the richer nations. □

John Cairns is at the Clinical Trial Service Unit, Harkness Building, Radcliffe Infirmary, Oxford OX2 6HE, UK.

Spin doctors

Paul Dirac: The Man and his Work

edited by Peter Goddard

Cambridge University Press: 1998. Pp. 124.

£12.95, \$19.95

The Story of Spin

by Sin-itiro Tomonaga, translated by

Takeshi Oka

University of Chicago Press: 1997. Pp. 258.

\$50, £39.95

Ian Aitchison

On the left, leaning elegantly backwards at an angle of 30 degrees to the vertical (but supported by solid stonework), head attentively inclined, is the slighter and older figure; on the right, vertically framing the space left by his listener's tilt, hands persuasively moulding the shape of the argument, is the handsome younger one. The old master and the young; the old world and the new; it is (as the caption says) "Dirac and Feynman discussing Physics".

The photograph, one of my favourite

images, is reproduced in Maurice Jacob's comprehensive lecture on antimatter, one of four given on 13 November 1995 to mark the dedication to Paul Dirac of a plaque in Westminster Abbey in London, and collected together in *Paul Dirac: The Man and his Work*, edited by Peter Goddard. The diamond-shaped plaque itself is also reproduced, opposite the first page of Stephen Hawking's Dirac memorial address, which introduces the volume.

The most remarkable thing about the plaque is that the Dirac equation is on it, the first equation (a latter-day commandment?) to reach the sanctum where many of Britain's cultural and artistic heroes are commemorated. It was surely a brilliant idea to include the equation on the plaque, the execution being helped, no doubt, by the use of Feynman's more compact gamma-matrix notation in place of Dirac's original alphas.

It may be some time before the compression of knowledge will allow well-educated visitors to enjoy its significance, but the Dirac equation will surely always continue to fascinate theoretical physicists. Which of them has not had the secret dream of being able, like Dirac, to discover a true and permanent part of the mathematical structure of the world after only a few sentences of simple argument? Yet, as Abraham Pais reminds us in his scholarly and intimate contribution, Dirac's own 'darling' was not the equation of 1928, but his transformation theory of 1927, which is both the simplest and most general mathematical formulation of quantum theory. As Pais says, such self-revelations by Dirac were rare, but all the more precious.

One such revelation I have always found

remarkable is quoted by David Olive in his lecture on monopoles. At the 1977 Enrico Fermi school in Varenna, Dirac confessed that the reason he had not, at a certain stage, pressed on with higher approximations (which would have tested his theory more severely) was that he "was afraid... the results might not come out right".

The collector of the more usual type of Dirac story will find many in the book. I much enjoyed one told by Jacob in which, after a lecture by Murray Gell-Mann on quarks, Dirac told Gell-Mann that he believed in quarks. "Wonderful," said Gell-Mann, overjoyed, "but what is your main reason?" "It is because they have spin," was Dirac's answer.

In arriving at his equation, which describes a relativistic particle of spin $1/2$, Dirac was famously motivated by considerations of mathematical beauty which, in his hands, could be a powerful heuristic tool in the search for physical laws. The fourth lecture in Goddard's collection is by Sir Michael Atiyah, and is the most mathematical (and succinct) of the four. Atiyah reviews the role in mathematics of the differential operator introduced by Dirac in his equation, which is essentially a formal square root of the wave operator. He then discusses various aspects of the Atiyah-Singer index theorem and its significance for physics in quantum violations of classical symmetries through 'anomalies'. Both Olive and Atiyah end their lectures by referring to recent fruitful interconnections between mathematics and physics in the work of Simon Donaldson and of Nathan Seiberg and Edward Witten. This is altogether a small gem of a book, which packs in a

great deal of information, anecdote and learning.

The Story of Spin is a refreshing contrast, though there are many fascinating connections. For example, in the first book, Pais mentions a trip to Japan taken by Dirac and Werner Heisenberg in 1929. Sin-itiro Tomonaga, then 23 years old and recently graduated from Kyoto University, went to Tokyo to hear their lectures. He reproduces a photograph of the occasion, and writes: "Miraculously, I could more or less understand the content of the lectures, because fortunately I had already looked through papers related to these talks... [although] this required great labour."

Tomonaga went on to share the 1965 Nobel prize for physics with Feynman and Julian Schwinger for the development of quantum electrodynamics, a theory initiated by Dirac. Unusually, his book, first published in Japanese in 1974, five years before his death, is devoted almost entirely to the single topic of spin, that strange attribute of particles which Wolfgang Pauli in 1924 called a "classically indescribable two-valuedness" and which Dirac's equation, in a sense, explains. But it was an inspired idea to give a course of lectures on spin, and in the process provide a fresh and personal perspective on the development of quantum mechanics from about 1920 to 1940. After all, as the translator Takeshi Oka remarks: "The existence of spin, and the statistics associated with it, is the most subtle and ingenious design of nature — without it the whole universe would collapse." This is, perhaps, part of what Dirac meant in his reply to Gell-Mann.

The spin-statistics connection was first proved by Pauli in 1940, and Tomonaga describes the argument in the eighth of his twelve lectures. The first three describe pre-quantal analysis of atomic spectra, the emergence of Pauli's two-valuedness idea, the triumphs and tragedies of the 'self-rotating electron' model, and then Pauli's spin theory and the Dirac equation. Next come the story of how proton spin was determined in 1927 from the specific heat of molecular hydrogen, and Heisenberg's explanation of ferromagnetism in terms of an exchange interaction and the spin-statistics connection. The sixth lecture describes the origins of field quantization, and the seventh gives an elementary account of the behaviour of that "mysterious tribe by the name of spinor family", as Paul Ehrenfest called these quantities, which are neither vector nor tensor.

By the ninth lecture we have reached 1932 and the discovery of the neutron (and its spin), followed by the "pulling down... of the walls of the nuclear sanctuary, which had been thought to be impenetrable by quantum mechanics." The story reaches its climax in the tenth lecture, with Heisenberg's theory of nuclear forces and the concept of isospin, Enrico Fermi's theory of beta-decay,



Past masters: Dirac (left) and Feynman.

and Hideki Yukawa's wonderful paper of 1935 in which he introduced the idea of virtual quanta, proposed the existence of a heavy nuclear force quantum, and prefigured today's Standard Model of beta-decay.

Tomonaga's style is informal and often humorous. In the epilogue, he thinks of "an old sailor talking on and on about his youthful adventures" as he "spins a yarn". The story of those 20-odd years in physics has been told before, but never from quite this perspective. Thanks to Oka's translation, Tomonaga's lectures will now be read and enjoyed by the wide audience they deserve. □

Ian Aitchison is in the Department of Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK.

Plant roots

The Origin and Early Diversification of Land Plants: A Cladistic Study

by Paul Kenrick and Peter R. Crane
Smithsonian Institution Press: 1997. Pp. 592.
\$55, £21.50

Sandy Knapp

The evolution and diversification of life has always held a great fascination for biologists. If we are to understand the patterns of relationship of life today, it is imperative that we have some understanding of those in the past; it puts our small branches in perspective. The extremes of innovation and diversification that have occurred throughout history tell compelling stories, but need to be put into clear systematic perspective to be analysed effectively.

For plants, the emergence of terrestrial forms was the start of a period of unparalleled diversification and innovation. The pivotal events leading to the origins and diversification of land plants are similar to those involved in the Cambrian 'explosion' that so influenced the diversification of the Metazoa. The metazoan fossils of the Burgess Shale endlessly fascinate both scientists and the general public, but the plant fossils of the Rhynie chert are just as fascinating from the evolutionary point of view.

The study of evolution depends on phylogeny: without hypotheses of the patterns of character distribution, and thus relationships between groups of organisms, no evolutionary hypotheses can really be constructed. The power and beauty of the approach of Paul Kenrick and Peter R. Crane to the analysis of land-plant relationships lies in their use of cladistic methodology. The focus on comparative morphology, and the testing of their own and other hypotheses of specific relationships, means that they truly achieve the objective set out in the first pages of the book: "to begin to synthesize the data that are currently available for resolving phylogenetic, and hence evolutionary, patterns



Colourful explanations

The island continent of Australia is known for its bizarre, colourful and often unique wildlife, from the marsupials and egg-laying mammals to the sea creatures of the Great Barrier Reef. Western Australia is particularly noted for its wildflowers, with more than 2,000 native species including the official state flower, the kangaroo

paw (above). In *A Natural History of Australia* (Academic Press, \$44.95), Tim M. Berra sets out to explain how the flora and fauna have been shaped by the isolation and aridity of the continent. The book is aimed at the interested visitor, incorporating a geography of Australia and useful information for the traveller.

among 'basal' groups of land plants."

This is not a book that tells a story based on an existing phylogeny; these are the real data, the characters that resolve the pattern, warts and all. It is tempting to think that evolution can be studied by simply plotting characters of interest on pre-existing phylogenies (usually constructed using molecular data), but Kenrick and Crane show how truly effective the alternative approach can be: that of an in-depth analysis of characters and their distribution using cladistic methodologies.

The book is constructed in a nesting set of detailed treatments of ever more detailed monophyletic groups. For example, the chapter on the relationships between embryophytes is followed by a chapter on the relationships among the polysporangio-phytes, tracheophytes and euphyllphytes, all members of the preceding group. Each chapter is constructed similarly, making the characters and methodology comparable at each stage of the analysis. The choice of taxa used in each analysis is detailed, and all characters used in the analysis are described in detail and their coding justified. This is unusual in a book of this sort, and is a welcome innovation. Kenrick and Crane also analyse carefully the degree of applicability of each character and state how many taxa had missing data for each character used.

This may seem like unnecessary detail for a synthetic book of this kind, but it allows one to assess the authors' confidence in a

character, rather than just taking their word for it. Characters with significant missing data — often those cellular details that are difficult to see in fossils — are also discussed in detail, and other potentially relevant characters not used in the analysis are considered as well. This honest and open approach, not often used because it reveals the limits of both the data and those analysing it, will surely stimulate palaeobotanists to re-examine fossils in museums around the world to redefine and refine the knowledge of these potentially useful characters.

The cladistic classification of the plant kingdom (Chlorobionta) is a model for how monophyly can be used as the defining principle for classification. A table of synapomorphy-based definitions of monophyletic higher taxa provides an extremely effective summary of the power of the cladistic approach. The in-depth discussion of developmental transformations in the context of taxic homology puts the characters used in Kenrick and Crane's analyses in perspective and allows them to set the stage for the next advances in the understanding of land-plant evolution.

This excellent and detailed book stands as a model for how to approach the study of evolution, and is an essential addition to the bookshelves of anyone interested in the scope and diversification of life. □

Sandy Knapp is in the Department of Botany, Natural History Museum, Cromwell Road, London SW7 5BD, UK.

Chat shows

Passionate Minds: The Inner World of Scientists

by Lewis Wolpert and Alison Richards
Oxford University Press: 1997. Pp. 240.
£19.99, \$25

A Glorious Accident: Understanding Our Place in the Cosmic Puzzle

by Wim Kayzer
W. H. Freeman: 1997. Pp. 306. \$24.95,
£17.95

Peter Atkins

Scratching a scientist reveals a human within. But how do humans who have made such extraordinary contributions to our understanding of the world differ from the rest of us? Lewis Wolpert set out, in the company of his producer Alison Richards, to expose the fire in the bellies of a clutch of substantial scientists. The television programmes they compiled are transcribed into two dozen short chapters in this engaging volume.

Their guests included Carl Djerassi, who with his synthesis of the first steroid contraceptive has probably done more for the liberation of humanity than any other chemist, Gerald Edelman, Murray Gell-Mann and James Black, whose work on receptor blockers has saved thousands of lives. Inevitably, I suppose, as they were seeking the established, reputed scientists of the late twentieth century, their guest list is late-middle-aged and dominated by men.

I was enthralled by the encapsulated biographies tweaked from their vessels by conversational nudges from Wolpert, who delicately and unobtrusively guides the conversations in directions that provide sometimes charming, sometimes deep insights into the springs of creativity. Determination and bloody-mindedness played a part in many of their lives, along with the ability to stand out against the flow of contemporary thought and ask questions that might be thought absurd but turned out to be amazingly fruitful. Of course, the raising of seemingly absurd questions is not on its own a sufficient condition for substantial discovery: if it were, all of us would have won Nobel prizes. Here we see seeming absurdity in the matrix of intelligence, application and talent.

I found a delight on every page, and warmly recommend this book to anyone with an interest in the human side of science, and not only that, for the speakers' own assessment of their achievements — for the most part modest, but throughout with a discernible and proper edge of pride — introduces us to their science in the gentlest possible way. I enjoyed travelling with Jared Diamond through New Guinea and listening to him robustly defending mere collectors of information (which he is not), and brushing out fossils with Elwyn Simons, as much as I enjoyed following Sheldon Glashow and Carlo Rubbia through the tangled history of the seminal contributions to our cosmic foundations that particle physics represents. In short, this book is a hugely enjoyable and instructive read.

The collection of interviews introduced and edited by Wim Kayzer also grew from another medium: they are a companion to a television series first shown in the Netherlands but also shown in the United States. Kayzer, whom Stephen Jay Gould so often brands as a romantic, had the idea of assembling some 'Great Brains' around a table, like a latter-day Ottoline Morrell, and watching what happened. That was the television component. To provide a background to the series, Kayzer ferreted around in the backgrounds of his contributors and, through a series of individual interviews that make up the bulk of this book, sought, like Wolpert in his, to bare the personal springs of public achievements.

The assembled Brains are those of Oliver Sacks, Stephen Jay Gould, Daniel Dennett, Freeman Dyson, Rupert Sheldrake and Stephen Toulmin. I think it regrettable that Sheldrake was invited, for any airing of absurd views adds to their life; but he is despatched with finesse by the orthodox team, and their general scepticism turns out to be a portrayal of real science in action. The remaining Brains span the cultures of the enterprise: Sacks erudite but whimsical, Gould acerbic but focused, Dennett imaginative and trenchant, Dyson precise and informed, and Toulmin cultured and circumspect.

All the interviews reveal the inner world of the participants, and in that respect Kayzer's book is an interesting complement to that of Wolpert and Richards. Two inter-

views stand out and leave a lasting impression. The first is Dyson's account of his migration from pacifism in the 1930s to being a contributing agent to countless deaths (and almost certainly the saviour of countless more) during the Second World War, and is moving for its honesty. I am not sure that it makes him 'human' in the generally accepted sense of the term, but it certainly adds a dimension to my understanding of this extraordinarily imaginative and talented man. The other outstanding interview is that with Gould, who plainly hated every second of it, and Kayzer had to drag the conversation out of him. The result is pure gold: pithy, acerbic and intelligent, with an occasional burst of words when exasperation got the upper hand and led to a torrent of truths being unleashed.

Gould maintains his role as guard of the rational and scourge of the soppy in the transcription of the discussion that follows. The discussion fulfilled Kayzer's original romantic dream of overhearing Brains at play in the sandpit of discussion. The Brains gossip eruditely and with spirit about some vaguely 'unanswerable' questions, such as the current 'Top Question' — the nature of consciousness — and some lesser, more readily answerable questions, such as what keeps the Brains spellbound today.

What keeps brains, as distinct from Brains, spellbound is of course the stuff of science, and these two books put such valuable insights into the process into the hands of the general reader. They are two thoroughly good and entertaining reads. □

Peter Atkins is at Lincoln College, Oxford OX1 3DR, UK.

New Journals

This year, *Nature's* annual new journals review supplement will appear in the issue of 10 September. Publishers and learned societies are invited to submit journals for review, as well as details of any eligible electronic journals, taking note of the following criteria:

- Journals must have first appeared during or after June 1996 and issued at least four separate numbers by the end of May 1998.
- Journals covering any aspect of science are eligible, although those dealing with clinical medicine and pure mathematics are excluded, as are newsletters and publications of abstracts.
- Frequency of publication must be at least three times a year.
- The main language is English.
- Deadline for submission is 5 June.

Please send at least four different issues (the first, the most recent and any two others) of each eligible title, together with full details of subscription rates, to: Peter Tallack, *Nature*, Porters South, Crinan Street, London N1 9XW, UK. Tel: +44 (0)171 843 4567. e-mail: p.tallack@nature.com

New in paperback

The Fabric of Reality

by David Deutsch
Penguin, £8.99
Reviewed in *Nature* 388, 136 (1997).

Brainchildren: Essays on Designing Minds

by Daniel C. Dennett
Penguin, £12.50

Collection of previously published essays by the garrulous Tufts University philosopher.

The Very First Light: A Scientific Journey Back to the Dawn of the Universe

by John C. Mather and John Boslough
Penguin, £8.99
Reviewed in *Nature* 385, 691 (1997).

Asymmetric cell division

Yuh Nung Jan & Lily Yeh Jan

With the recent identification of intrinsic cell-fate determinants for asymmetric cell division in several systems, biologists have begun to gain insight into the cellular mechanisms by which these determinants are preferentially segregated into one of the two daughter cells during mitosis so that the daughter cells acquire different fates.

Asymmetric cell division, in which a cell divides into two cells of different developmental potentials, is a fundamental means of generating cell diversity. In principle, asymmetric cell divisions may involve extrinsic or intrinsic factors. With extrinsic factors, daughter cells are initially equivalent but adopt different fates as the result of the interactions of the daughter cells with each other or with their environment. With intrinsic factors, unequal amounts of cell-fate determinants are partitioned into the two daughter cells. The mechanisms underlying asymmetric cell division have been studied in organisms ranging from bacteria, yeast, worms (*Caenorhabditis elegans*) and flies (*Drosophila*) to mammals^{1–6}. Intrinsic determinants have been identified in several systems, raising some key questions in molecular terms. What controls the asymmetric localization of the determinants during mitosis? How do the determinants control daughter cell fate? Here we discuss these topics, focusing primarily on the *Drosophila* nervous system and using other systems for comparison.

Identification of the intrinsic cell-fate determinants

The unicellular budding yeast divides with a characteristic polarity and asymmetry of cell fate. A smaller 'daughter' cell buds off from the larger 'mother' cell. The mother cell can switch mating type but the daughter cannot. This asymmetry exists because the mother, but not the daughter, cell expresses the *HO* endonuclease which catalyses the genetic recombination event that leads to mating-type switching. The *Ash1* protein is an intrinsic determinant for this asymmetric division^{7–9}. *Ash1* is normally found only in the daughter cell and it is a nuclear protein that functions as a transcriptional repressor of *HO* (Fig. 1a). In loss-of-function *ASH1* mutants, both the mother cell and the daughter cell can switch mating type^{8,9}. Conversely, overexpression of *ASH1* in mother cell as well as daughter cell prevents mating-type switching in both cells⁹. *Ash1* expression is cell-cycle regulated; *ASH1* is transcribed during mitosis, *Ash1* messenger RNA is localized to the distal tip of the budding daughter cell in post-anaphase, and *Ash1* is present in the daughter nucleus in late anaphase and early G1 phase^{9–11}.

Asymmetric divisions generate the different cell types of a multicellular organism. During the development of the fly peripheral nervous system (PNS), the four cells of a sensory bristle, the two outer cells (hair and socket cells) and two inner cells (neuron and sheath cell), arise from a single sensory organ precursor (SOP) from two rounds of asymmetric cell division. In the first round of division, in wild-type flies, the SOP divides into a Ila cell (the precursor of hair and socket) and a I Ib cell (the precursor of neuron and sheath cell). In *numb* loss-of-function mutants, the SOP divides symmetrically and gives rise to two Ila cells, which produce outer cells but not neurons¹².

Numb is a membrane-associated protein and is asymmetrically localized in a cell-cycle-dependent manner^{13,14}. In a dividing SOP, *Numb* is initially uniformly distributed. During late prophase and metaphase, *Numb* forms a crescent overlying one of the two spindle poles, and then segregates predominantly into the I Ib cell upon division of the SOP (Fig. 1b)^{13,14}. The amount of *Numb* determines the cell fate. In loss-of-function *numb* mutants, both daughters of

SOP become Ila cells¹². Conversely, if *Numb* is overexpressed in the SOP, then both daughters develop into I Ib cells¹³. These results show that *Numb* functions as an intrinsic cell-fate determinant, because its preferential segregation into one of the two daughter cells causes that cell to adopt the I Ib fate¹³.

Numb is also asymmetrically localized in neuroblasts, the neural precursors in the central nervous system (CNS). As a neuroblast divides to give rise to another neuroblast and a ganglion mother cell (GMC), which divides only once to produce two neurons, *Numb* is segregated preferentially into the GMC^{13,14}. *Numb* is required for the asymmetric cell division of at least some CNS lineages¹⁵.

Genes in asymmetric cell division in *Drosophila*

In principle, there are several distinct (but not mutually exclusive) means by which a protein such as *Numb* could be asymmetrically localized in a dividing precursor cell so as to be preferentially segregated into one of the two progeny cells. For example, the asymmetric localization of this protein could be the result of asymmetric localization of the mRNA that encodes the protein. The protein could also be localized asymmetrically by a certain transport system. It is also possible that the protein is selectively degraded in a portion of the cell, leaving the remaining protein to form a crescent. In order to gain insight about the underlying mechanisms, genetic and biochemical studies have established a rudimentary genetic pathway for asymmetric localization of *Numb* and other factors (Fig. 2). Of the genes identified so far, *inscuteable* is at the top of the hierarchy^{16,17}. It coordinates three aspects of asymmetric cell divisions in the nervous system as follows.

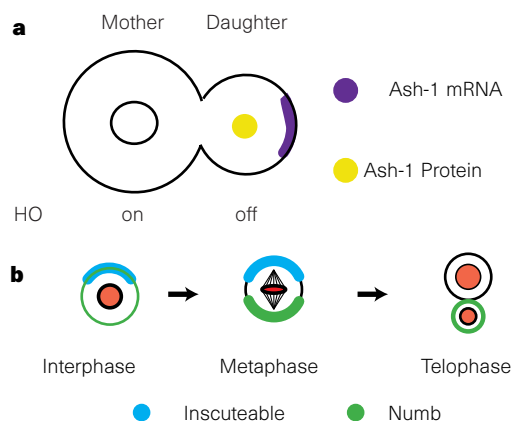


Figure 1 Asymmetric localization of determinants. **a**, The budding yeast divides asymmetrically to produce a mother cell and a smaller daughter cell. Asymmetric localization of *Ash1* mRNA causes the daughter cell to inherit the *Ash1* protein, a negative transcription regulator that turns off *HO* expression. **b**, In the fruit fly, a neural precursor cell divides asymmetrically to produce two different daughter cells. *Inscuteable* first forms a crescent at one side of the neural precursor cell during interphase. *Inscuteable* is required for a cell-fate determinant, *Numb*, to form a crescent at the opposite side during metaphase. After division, one of the daughter cells inherits *Numb* and assumes a fate distinct from its sibling.

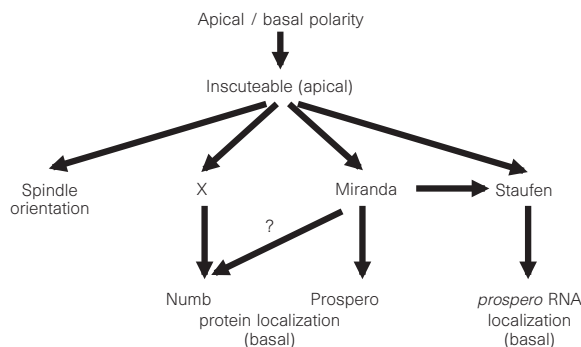


Figure 2 A genetic pathway for asymmetric localization of determinants in *Drosophila*. The apical/basal polarity is laid down early during embryogenesis and somehow directs the apical localization of Inscuteable in the neuroblast. This asymmetric localization of Inscuteable is necessary for spindle orientation along the apical/basal axis as well as asymmetric localization of Numb. Prospero and *prospero* mRNA are localized to the basal side of the neuroblast. Asymmetric localization of Prospero and its mRNA depends on Miranda and Staufien.

(1) Asymmetric localization of Numb, Prospero and Miranda.

Prospero and Miranda, like Numb, show cell-cycle-dependent asymmetric localization in dividing neural precursors in the CNS and the PNS. Prospero is a homeobox-containing transcriptional regulator that is required for proper neuronal differentiation, neuronal cell-fate specification and axon growth and guidance^{18,19}. Whereas Prospero is normally a nuclear protein, it translocates to the cell membrane after nuclear envelope breakdown during mitosis. By metaphase, the Prospero protein forms a crescent and is colocalized with Numb. After cell division, both Numb and Prospero are segregated into the same daughter cell with Numb staying at the membrane and Prospero entering the nucleus^{14,20,21}. Although Numb and Prospero are colocalized during part of the mitotic cycle, their localizations are independent of each other; the Prospero crescent forms normally in *numb* mutants and the Numb crescent is normal in *prospero* mutants^{14,21}. Because the Numb protein appears to accumulate as the crescent forms, it seems more likely that protein translocation rather than selective degradation is the primary cause of Numb asymmetric localization. However, one cannot exclude the possibility that selective degradation plays a role.

Asymmetric localization of Prospero requires Miranda, a new membrane-associated protein for which the coiled-coil structure has been predicted^{22,23}. In dividing neural precursors, Miranda also forms a crescent and colocalizes with Numb and Prospero during metaphase. Miranda can bind Prospero and appears to be an adaptor protein required to bring Prospero to the cell membrane to form a crescent. In *miranda* loss-of-function mutants, Prospero is distributed uniformly in the cytoplasm and fails to reach the cell membrane^{22,23}.

The proper localization of Numb, Prospero and Miranda requires Inscuteable^{17,22}. However, this requirement varies to some extent in different parts of the nervous system. In the neuroblasts of *inscuteable* mutants, Numb, Prospero and Miranda are either delocalized or form crescents that are no longer localized to the basal cortex but exhibit random orientation^{17,22}. In contrast, the *inscuteable* mutation has much weaker effects in the PNS^{16,17}, suggesting that additional factors are involved.

(2) **Asymmetric RNA localization.** The mRNAs that encode some of the asymmetrically localized proteins are also asymmetrically localized in the fly neuroblasts^{16,24}. *prospero* mRNA is asymmetrically localized to the apical cell cortex in the interphase neuroblast and relocates to the basal cell cortex during mitosis²⁴. This RNA localization requires the Staufien protein²⁴. Staufien encodes an RNA-binding protein that is part of the machinery required for the localization of the anterior determinant *bicoid* and the posterior

determinant *nanos* during fly oogenesis²⁵. Staufien is the first link found between asymmetric cell division in the nervous system and the asymmetric localization of determinants involved in anteroposterior patterning during oogenesis. It will be interesting to determine whether additional components are shared between these two systems. In *staufien* mutants, *prospero* mRNA fails to localize to the basal surface²⁴. However, the Prospero protein is still asymmetrically localized to the basal cell cortex. Moreover, *staufien* mutants can develop into viable adults, whereas *numb* and *prospero* mutants die as embryos. It could be that *prospero* mRNA localization serves as a redundant or back-up system for Prospero localization. Alternatively, *prospero* mRNA localization may serve an unknown function.

In contrast to the *prospero* mRNA localization being dispensable for Prospero asymmetric localization, the asymmetric Ash1 mRNA localization is a prerequisite for the asymmetric location of Ash1 to the daughter nucleus^{10,11}. As in many cases of mRNA localization, the Ash1 mRNA localization is at least in part mediated through its 3'-UTR (untranslated region)²⁶.

(3) **Mitotic spindle orientation.** During fly neurogenesis, the neuroblasts delaminate from a monolayer of ectodermal cells. The ectodermal cells at the surface of the *Drosophila* embryo divide with the axes of their mitotic spindle parallel to the plane of the ectodermal monolayer. In contrast, a neuroblast divides along the apical-basal axis, so that the axes of its mitotic spindles are perpendicular to the plane of ectodermal layer. Thus, the mitotic spindle of the neuroblast has to reorientate by 90 degrees from the plane of the ectodermal layer. The rotation requires Inscuteable¹⁷. In *inscuteable* loss-of-function mutants, the orientation of the mitotic spindle in dividing neuroblasts becomes random. Conversely, when *inscuteable* is misexpressed in the ectodermal cells (which normally do not express *inscuteable*), those cells reorientate their mitotic spindles and divide along the apical-basal axis¹⁷.

The orientation of the mitotic spindle correlates with the location of the Numb, Prospero and Miranda crescents during mitosis of neuroblasts. The three proteins are colocalized in a crescent overlying the basal spindle pole. This ensures that the basal daughter will inherit most of these proteins (Fig. 1b). Both spindle orientation and asymmetric protein localization are under the control of Inscuteable¹⁷. However, they appear to be controlled independently: when spindle orientation is disrupted by microtubule depolymerizing agents, these proteins still form a basal crescent.

How might Inscuteable function? Inscuteable is asymmetrically localized to the apical pole of dividing neural precursors before they enter prophase^{16,17}. As the cell cycle progresses, Miranda, Prospero and Staufien become transiently localized to the apical pole^{21,24}, raising the possibility that Inscuteable may interact with these proteins directly or indirectly and, through an unknown mechanism, localize these proteins to the basal pole. By metaphase, Miranda, Staufien, Prospero and Numb form a basal crescent that is well separated from the apical crescent of Inscuteable (Fig. 2).

Inscuteable is a new protein with a central portion containing several repeats that bear moderate resemblance to ankyrin repeats¹⁶. This central portion is responsible for Inscuteable asymmetric localization as well as its ability to reorientate the mitotic spindle and to localize Numb and Prospero (J. A. Knoblich *et al.*, manuscript in preparation). Inscuteable localizes Numb, Prospero and *prospero* mRNA through intermediaries (Fig. 2). Miranda, a multi-domain adaptor that is capable of binding Inscuteable, Prospero and Staufien, is such an intermediary that is required for the asymmetric localization of Prospero, Staufien and *prospero* mRNA⁵¹.

Localization may depend on cytoskeletal elements

For asymmetric localization of proteins in fly neural precursors, the proteins may have to be first localized to the membrane either directly or through an adaptor. In the absence of membrane localization, as in the case of Prospero in *miranda* mutants^{22,23} or the Numb or Inscuteable protein with its membrane-localization

domain deleted (ref. 27 and J. A. Knoblich *et al.*, manuscript in preparation), the proteins stay in the cytoplasm and fail to be asymmetrically localized. The membrane association of these proteins before their asymmetric localization may be an example of 'reduction of dimensionality'²⁸: the time it takes for a molecule to reach a particular location can be greatly reduced by restricting its diffusion to two-dimensional instead of three-dimensional space.

The asymmetric localization of these proteins requires certain cytoskeletal elements. In embryos treated with the microfilament inhibitor cytochalasin D, *Inscuteable* is no longer asymmetrically localized¹⁷. Although Numb and Prospero are still asymmetrically localized in neuroblasts, the crescents frequently form at incorrect positions^{14,17}. Treatment with the more potent microfilament inhibitor latrunculin A abolishes Numb and Prospero asymmetric localization^{27,29}. It thus appears that the asymmetric localization of *Inscuteable* differs from Numb/Prospero asymmetric localization in the requirements for intact microfilaments. The involvement of microfilaments in the asymmetric localization of determinants seems to be a general requirement. In addition to a requirement for microfilaments, the localization of Ash1 in budding yeast, P-granules in *C. elegans* and posterior determinants during *Drosophila* oogenesis also involves Myo4 (refs 7, 8), non muscle myosin II³⁰, and tropomyosin^{31,32}, respectively, suggesting that motor-based transport systems are utilized.

Compared to microfilaments, the requirement of microtubules is more variable. Ash1 localization^{10,11} and the localization of *Inscuteable*, Numb and Prospero do not require microtubules^{14,17}. In contrast, the localization of both anterior and posterior determinants during fly oogenesis requires microtubules^{33,34}.

What might be the mechanisms for asymmetric localization of proteins such as Numb? There are several possible and not mutually exclusive models (for a more detailed discussion, see ref. 35). (1) The transport model: Numb is bound to a motor and carried along actin or another cytoskeletal structure toward its target area. This may be analogous to the potential involvement of an actin/myosin-based transport system for the transport of Ash1 mRNA in budding yeast^{7,8} and the asymmetric localization of PAR-1, 2 and 3 in *C. elegans*³⁰. (2) The diffusion model: Numb can diffuse along the cell membrane and is then captured by a localized anchor. (3) The capping model: the Numb proteins form aggregates either with themselves or with other molecules and eventually form a cap on one side of the cell. We do not know whether one or a combination of these models accounts for Numb localization.

How do the determinants control cell fates?

Determinants such as Ash1 and Prospero are transcriptional regulators. Unequal segregation of those determinants between the daughter cells results in differential gene expression. Ash1 suppresses *HO* transcription and thereby controls mating-type switching^{8,9}. Prospero can be either a positive or a negative regulator depending on its cofactor³⁶. It may control the expression of genes involved in neuronal differentiation^{18,19}, such as *deadpan*, *asense*, *even-skipped* and *fushi-tarazu*.

In contrast to nuclear proteins that regulate transcription, Numb functions by modulating cell–cell interactions. Cell–cell interaction mediated by the transmembrane receptor Notch is required for the four progeny of an SOP to assume their correct fate. Without Notch activity, all four cells become neurons³⁷. During each cell division within the sensory organ lineage, Numb appears to bias the Notch-mediated cell–cell interaction by inhibiting Notch activity^{38,39}, so that the cell–cell interaction becomes asymmetric. The daughter that inherits Numb has lower Notch activity relative to its sibling and adopts IIB fate (Fig. 3). This is an example of how an intrinsic mechanism using Numb and an extrinsic mechanism mediated by Notch can be integrated for the control of cell fate.

How might Numb inhibit Notch function? Numb is a membrane protein containing a phosphotyrosine-binding domain. Numb can

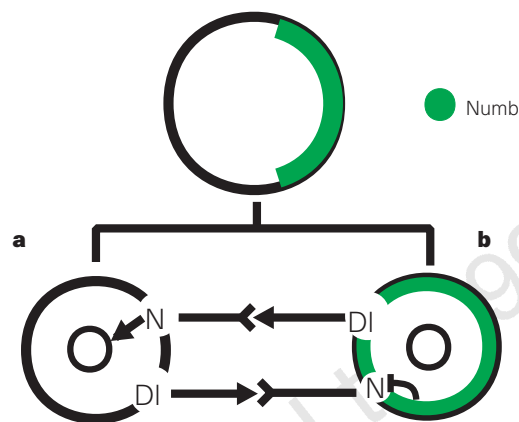


Figure 3 An interplay between the intrinsic and extrinsic mechanisms for asymmetric cell division. The intrinsic determinant Numb inhibits Notch activity in the interaction between the two daughter cells. The Delta (DI) ligand and Notch (N) receptor are expressed by both daughter cells and mediate the cell–cell interaction between them. The daughter cell that inherits Numb (**b**) has its Notch activity inhibited, so that it receives weaker Notch signalling than its sister cell (**a**).

bind Notch in *in vitro* binding assays³⁸. However, the mechanism by which Numb inhibits Notch is unknown. Numb could inhibit Notch activity by direct interaction. Alternatively, Numb could function as an adaptor protein to recruit other proteins that regulate Notch signalling. This inhibition of Notch by Numb is used as a general binary switch rather than for the specification of a particular cell fate. Besides PNS, it is also used in specifying two alternative sibling fates in certain CNS lineages^{15,40} and muscle founder cell lineages in *Drosophila* (ref. 41 and R. Bodmer and W. Chia, personal communication). In different lineages, different downstream effectors may be used to specify the daughter cell fates. An example is the sensory organ lineage in which the asymmetric divisions of both IIA and IIB employ the Numb/Notch binary switch; however, *Su(H)* is used to transduce Notch signalling in IIA but not in IIB lineages⁴².

C. elegans has an alternative way of making Notch signalling asymmetric. During the 4–8-cell stage of *C. elegans* embryogenesis, the anterior cells receive inductive signals from the posterior cells. This signal is mediated by GLP-1 (a *C. elegans* Notch homologue). The asymmetry of GLP-1 signalling is at least in part achieved by selective translation in the anterior cells of *glp-1* mRNA, which is uniformly distributed in both anterior and posterior cells⁴³.

Evolutionary considerations

The molecules that mediate cell–cell interaction during asymmetric cell division appear to be well conserved, including the molecules in the Notch signalling pathway and the Wnt signalling pathway^{44–46}. By contrast, the molecules that have been found thus far to be involved in asymmetric localization of determinants in *Drosophila*, *C. elegans*^{4–6} or budding yeast^{7,8} do not bear any family resemblance, except for the involvement of Staufin in both neural development and oogenesis in *Drosophila*. This apparent diversity of molecular mechanisms may simply reflect our current state of ignorance. Once the remaining components are identified in those systems, a common theme may emerge. Alternatively, the diversity may be genuine and reflect the different strategies evolved for achieving asymmetric localization of determinants in different situations. For example, the distance to be covered by determinants in an oocyte (up to hundreds of micrometres) is much greater than that in a neural precursor (less than 10 μ m). Furthermore, the anterior and posterior determinants are laid down only once during oogenesis whereas determinants for asymmetric division are asymmetrically localized every time the neural precursor divides.

Homologues for a few of the genes such as *numb* and *par-1* have been found in vertebrates. The mouse Numb, when introduced into

the fly, is asymmetrically localized and can rescue the fly *numb* mutant phenotype⁴⁷. Mouse Numb (mNumb) is also asymmetrically localized in the neural precursor cells in developing mouse cortex; however, there is an interesting difference. In mouse, mNumb crescent localization and spindle orientation are not coupled as in the fly. Whereas a fly neural precursor always divides asymmetrically and imparts Numb to one of its daughters, a mouse neural precursor may divide either asymmetrically, with one daughter inheriting most of the mNumb, or symmetrically, with both daughters inheriting mNumb⁴⁷. The *in vivo* function of mNumb remains to be determined. PAR-1 is a putative Ser/Thr kinase required for the first asymmetric cell division of a *C. elegans* zygote⁴⁸. The mammalian PAR-1 homologues, mPAR-1 and MARK, are kinases that can phosphorylate microtubule-associated proteins^{49,50}. Further, their overexpression disrupts the microtubule cytoskeleton and alters cell shape. These results suggest that PAR-1 may have a conserved function in specifying cell polarity by regulating the microtubule cytoskeleton.

Perspective and future directions

Recent studies have revealed multistep machineries that localize intrinsic determinants for asymmetric cell division. Our understanding of the underlying mechanisms is still fairly rudimentary. Besides the obvious need to identify the remaining components of the machineries, questions that need to be addressed include the following. (1) Inscuteable does not set up cell polarity. Instead, it probably responds to the apical-basal polarity information that is laid down early in *Drosophila* development (perhaps at cellular blastoderm stage). How does Inscuteable transduce this apical-basal polarity information? (2) How is the apical-basal polarity used to orientate the mitotic spindle? (3) The asymmetric localization of determinants is tightly linked with the cell cycle. How does the basic cell cycle machinery regulate the asymmetric localization of determinants? Is it through cell-cycle-dependent post-translational modification and/or degradation of the localization machinery? (4) The asymmetry organizer appears to have two centres at opposite poles of the cell, for example an apical centre for Inscuteable localization and a basal centre for the Numb/Prospero/Miranda crescent in fly neuroblasts (Fig. 1b). How do the two centres communicate with each other? (5) Are the intrinsic determinants translocated by transport, diffusion or capping? (6) The Numb/Notch interaction provides one way of integrating intrinsic and extrinsic mechanisms. How are the other signalling pathways such as the Wnt pathway integrated with the intrinsic determinants' functions? (7) Have different organisms adopted different strategies, or are there common mechanisms for asymmetric cell divisions? The intensive efforts being directed at several experimental systems should soon deliver some answers. □

Yuh Nung Jan and Lily Yeh Jan are at the Howard Hughes Medical Institute, Departments of Physiology, Biochemistry & Biophysics, University of California at San Francisco, Parnassus and Third Avenues, San Francisco, California 94143-0725, USA.

- Horvitz, H. R. & Herskowitz, I. Mechanisms of asymmetric cell division: two Bs or not two Bs, that is the question. *Cell* **68**, 237–255 (1992).
- Shapiro, L. & Losick, R. Protein localization and cell fate in bacteria. *Science* **276**, 712–718 (1997).
- Chang, F. & Drubin, D. G. Cell division: why daughters cannot be like their mothers. *Curr. Biol.* **6**, 651–654 (1996).
- Guo, S. & Kemphues, K. J. Molecular genetics of asymmetric cleavage in the early *Caenorhabditis elegans* embryo. *Curr. Opin. Genet. Dev.* **6**, 408–415 (1996).
- Kemphues, K. J. & Strome, S. in *C. elegans II* (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 335–359 (Cold Spring Harbor Laboratory Press, New York, 1997).
- Schnabel, R. & Priess, J. R. in *C. elegans II* (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 361–382 (Cold Spring Harbor Laboratory Press, New York, 1997).
- Jansen, R. P., Dowzer, C., Michaelis, C., Galova, M. & Nasmyth, K. Mother cell-specific *HO* expression in budding yeast depends on the unconventional myosin Myo4p and other cytoplasmic proteins. *Cell* **84**, 687–697 (1996).
- Bobola, N., Jansen, R. P., Shin, T. H. & Nasmyth, K. Asymmetric accumulation of Ash1p in postanaphase nuclei depends on a myosin and restricts yeast mating-type switching to mother cells. *Cell* **84**, 699–709 (1996).
- Sil, A. & Herskowitz, I. Identification of an asymmetrically localized determinant, Ash1p, required for lineage-specific transcription of the yeast *HO* gene. *Cell* **84**, 711–722 (1996).
- Long, R. M. *et al.* Mating type switching in yeast controlled by asymmetric localization of *ASH1*

- mRNA. *Science* **277**, 383–387 (1997).
- Takizawa, P. A., Sil, A., Swedlow, J. R., Herskowitz, I. & Vale, R. D. Actin-dependent localization of an RNA encoding a cell-fate determinant in yeast. *Nature* **389**, 90–93 (1997).
- Uemura, T., Shepherd, S., Ackerman, L., Jan, L. Y. & Jan, Y. N. *numb*, a gene required in determination of cell fate during sensory organ formation in *Drosophila* embryo. *Cell* **58**, 349–360 (1989).
- Rhyu, M. S., Jan, L. Y. & Jan, Y. N. Asymmetric distribution of Numb protein during division of the Sensory Organ Precursor cell confers distinct fates to daughter cells. *Cell* **76**, 477–491 (1994).
- Knoblich, J. A., Jan, L. Y. & Jan, Y. N. Asymmetric segregation of Numb and Prospero during cell division. *Nature* **377**, 624–627 (1995).
- Spana, E. P. *et al.* Asymmetric localization of Numb autonomously determines sibling neuron identity in *Drosophila* CNS. *Development* **121**, 3489–3494 (1995).
- Kraut, R. & Campos-Ortega, J. A. *inscuteable*, a neural precursor gene of *Drosophila*, encodes a candidate for a cytoskeleton adapter protein. *Dev. Biol.* **174**, 65–81 (1996).
- Kraut, R., Chia, W., Jan, L. Y., Jan, Y. N. & Knoblich, J. A. Role of *inscuteable* in orienting asymmetric cell divisions in *Drosophila*. *Nature* **383**, 50–55 (1996).
- Doe, C. Q., Chu-LaGriff, Q., Wright, D. M. & Scott, M. P. The *prospero* gene specifies cell fates in the *Drosophila* central nervous system. *Cell* **65**, 451–465 (1991).
- Vaessin, H. *et al.* *prospero* is expressed in neuronal precursors and encodes a nuclear protein that is involved in the control of axonal outgrowth in *Drosophila*. *Cell* **67**, 941–953 (1991).
- Hirata, J., Nakagoshi, H., Nabeshima, Y. & Matsuzaki, F. Asymmetric segregation of a homeoprotein, Prospero, during cell divisions in neural and endodermal development. *Nature* **377**, 627–630 (1995).
- Spana, E. P. & Doe, C. Q. The Prospero transcription factor is asymmetrically localized to the cell cortex during neuroblast mitosis in *Drosophila*. *Development* **121**, 3187–3195 (1995).
- Shen, C. P., Jan, L. Y. & Jan, Y. N. Miranda is required for the asymmetric localization of Prospero during mitosis in *Drosophila*. *Cell* **90**, 449–458 (1997).
- Ikehashima-Kataoka, H., Skeath, J. B., Nabeshima, Y., Doe, C. Q. & Matsuzaki, F. Miranda directs Prospero to a daughter cell during *Drosophila* asymmetric divisions. *Nature* **390**, 625–629 (1997).
- Li, P. *et al.* Inscuteable and Staufin mediate asymmetric localization and segregation of *prospero* RNA during *Drosophila* neuroblast cell divisions. *Cell* **90**, 437–447 (1997).
- St. Johnston, D., Beuchle, D. & Nüsslein-Volhard, C. *staufer*, a gene required to localize maternal RNAs in the *Drosophila* egg. *Cell* **66**, 51–63 (1991).
- Grünert, S. & St. Johnston, D. RNA localization and the development of asymmetry during *Drosophila* oogenesis. *Curr. Opin. Genet. Dev.* **6**, 395–402 (1996).
- Knoblich, J. A., Jan, L. Y. & Jan, Y. N. The N terminus of the *Drosophila* Numb protein directs membrane association and actin-dependent asymmetric localization. *Proc. Natl Acad. Sci. USA* **94**, 13005–13010 (1997).
- Adam, G. & Delbrück, M. in *Structural Chemistry and Molecular Biology* (eds Rich, A. & Davidson, N.) 198–215 (Freeman, San Francisco, 1968).
- Broadus, J. & Doe, C. Q. Extrinsic cues, intrinsic cues and microfilaments regulate asymmetric protein localization in *Drosophila* neuroblasts. *Curr. Biol.* **7**, 827–835 (1997).
- Guo, S. & Kemphues, K. J. A non-muscle myosin required for embryonic polarity in *Caenorhabditis elegans*. *Nature* **382**, 455–458 (1996).
- Erdelyi, M., Michon, A.-M., Guichet, A., Glotzer, J. B. & Ephrussi, A. Requirement for *Drosophila* cytoplasmic tropomyosin in *oskar* mRNA localization. *Nature* **377**, 524–527 (1995).
- Tetzlaff, M. T., Jackle, H. & Pankratz, M. J. Lack of *Drosophila* cytoskeletal tropomyosin affects head morphogenesis and the accumulation of *oskar* mRNA required for germ cell formation. *EMBO J.* **15**, 1247–1254 (1996).
- Clark, I. *et al.* Transient posterior localization of a kinesin fusion protein reflects anteroposterior polarity of the *Drosophila* oocyte. *Curr. Biol.* **4**, 289–300 (1994).
- Pokrywka, N. J. & Stephenson, E. C. Microtubules are a general component of mRNA localization systems in *Drosophila* oocytes. *Dev. Biol.* **167**, 363–370 (1995).
- Knoblich, J. A., Jan, L. Y. & Jan, Y. N. in *Cold Spring Harb. Symp. Quant. Biol.* **62**, 71–77 (1997).
- Hassan, B., Li, L., Bremer, K. A., Chang, W., Pinsonneault, J. & Vaessin, H. Prospero is a pan-neural transcriptional factor that modulates homeodomain protein activity. *Proc. Natl Acad. Sci. USA* **94**, 10991–10996 (1997).
- Hartenstein, V. & Posakony, J. W. A dual function of the *Notch* gene in *Drosophila sensillum* development. *Dev. Biol.* **142**, 13–30 (1990).
- Guo, M., Jan, L. Y. & Jan, Y. N. Control of daughter cell fates during asymmetric division; interaction of Numb and Notch. *Neuron* **17**, 27–41 (1996).
- Frise, E., Knoblich, J. A., Younger-Shepherd, S., Jan, L. Y. & Jan, Y. N. The *Drosophila* Numb protein inhibits signaling of the Notch receptor during cell-cell interaction in sensory organ lineage. *Proc. Natl Acad. Sci. USA* **93**, 11925–11932 (1996).
- Spana, E. P. & Doe, C. Q. Numb antagonizes Notch signaling to specify sibling neuron cell fates. *Neuron* **17**, 21–26 (1996).
- Gomez, R. & Bate, M. Segregation of myogenic lineage in *Drosophila* requires Numb. *Development* **124**, 4857–4866 (1997).
- Wang, S., Younger-Shepherd, S., Jan, L. Y. & Jan, Y. N. Only a subset of the binary cell fate decisions mediated by Numb/Notch signaling in *Drosophila* sensory organ lineage requires *Suppressor of Hairless*. *Development* **124**, 4435–4446 (1997).
- Evans, T. C., Crittenden, S. L., Kodoyianni, V. & Kimble, J. Translational control of maternal *gfp-1* mRNA establishes an asymmetry in the *C. elegans* embryo. *Cell* **77**, 183–194 (1994).
- Sawa, H., Lobel, L. & Horvitz, H. R. The *Caenorhabditis elegans* gene *lin-17*, which is required for certain asymmetric cell divisions, encodes a putative seven-transmembrane protein similar to the *Drosophila* Frizzled protein. *Genes Dev.* **10**, 2189–2197 (1996).
- Thorpe, C. J., Schlesinger, A., Clayton Cater, J. & Bowerman, B. Wnt signaling polarizes an early *C. elegans* blastomere to distinguish endoderm from mesoderm. *Cell* **90**, 695–705 (1997).
- Rocheleau, C. E. *et al.* Wnt signaling and an APC-related gene specify endoderm in early *C. elegans* embryos. *Cell* **90**, 707–716 (1997).
- Zhong, W., Feder, J. N., Jiang, M. M., Jan, L. Y. & Jan, Y. N. Asymmetric localization of a mammalian Numb homologue during mouse cortical neurogenesis. *Neuron* **17**, 43–53 (1996).
- Guo, S. & Kemphues, K. J. *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81**, 611–620 (1995).
- Böhm, H., Brinkmann, V., Drab, M., Henske, A. & Kurzchalia, T. V. Mammalian homologues of *C. elegans* PAR-1 are asymmetrically localized in epithelial cells and may influence their polarity. *Curr. Biol.* **7**, 603–606 (1997).
- Drewe, G., Ebner, A., Preuss, U., Mandelkow, E.-M. & Mandelkow, E. MARK, a novel family of protein kinases that phosphorylate microtubule-associated proteins and trigger microtubule disassembly. *Cell* **89**, 297–308 (1997).
- Shen, C.-P. *et al.* Miranda as a multidomain adaptor linking apically localized Inscuteable and basally localized Staufin and Prospero during asymmetric cell division in *Drosophila*. *Genes Dev.* (in the press).

Acknowledgements. We thank our colleagues S. Abdelilah, J. Brennan, F.-B. Gao, L. Sharp, C.-P. Shen, S. Younger-Shepherd, C. Zeng and especially Y.-M. Chan for helpful comments on this article. We are both Howard Hughes Medical Institute Investigators.

Global-scale temperature patterns and climate forcing over the past six centuries

Michael E. Mann*, Raymond S. Bradley* & Malcolm K. Hughes†

* Department of Geosciences, University of Massachusetts, Amherst, Massachusetts 01003-5820, USA

† Laboratory of Tree Ring Research, University of Arizona, Tucson, Arizona 85721, USA

Spatially resolved global reconstructions of annual surface temperature patterns over the past six centuries are based on the multivariate calibration of widely distributed high-resolution proxy climate indicators. Time-dependent correlations of the reconstructions with time-series records representing changes in greenhouse-gas concentrations, solar irradiance, and volcanic aerosols suggest that each of these factors has contributed to the climate variability of the past 400 years, with greenhouse gases emerging as the dominant forcing during the twentieth century. Northern Hemisphere mean annual temperatures for three of the past eight years are warmer than any other year since (at least) AD 1400.

Knowing both the spatial and temporal patterns of climate change over the past several centuries remains a key to assessing a possible anthropogenic impact on post-industrial climate¹. In addition to the possibility of warming due to increased concentrations of greenhouse gases during the past century, there is evidence that both solar irradiance and explosive volcanism have played an important part in forcing climate variations over the past several centuries^{2,3}. The unforced 'natural variability' of the climate system may also be quite important on multidecadal and century timescales^{4,5}. If a faithful empirical description of climate variability could be obtained for the past several centuries, a more confident estimation could be made of the roles of different external forcings and internal sources of variability on past and recent climate. Because widespread instrumental climate data are available for only about one century, we must use proxy climate indicators combined with any very long instrumental records that are available to obtain such an empirical description of large-scale climate variability during past centuries. A variety of studies have sought to use a 'multiproxy' approach to understand long-term climate variations, by analysing a widely distributed set of proxy and instrumental climate indicators^{1,5–8} to yield insights into long-term global climate variations. Building on such past studies, we take a new statistical approach to reconstructing global patterns of annual temperature back to the beginning of the fifteenth century, based on the calibration of multiproxy data networks by the dominant patterns of temperature variability in the instrumental record.

Using these statistically verifiable yearly global temperature reconstructions, we analyse the spatiotemporal patterns of climate change over the past 500 years, and then take an empirical approach to estimating the relationship between global temperature changes, variations in volcanic aerosols, solar irradiance and greenhouse-gas concentrations during the same period.

Data

We use a multiproxy network consisting of widely distributed high-quality annual-resolution proxy climate indicators, individually collected and formerly analysed by many palaeoclimate researchers (details and references are available: see Supplementary Information). The network includes (Fig. 1a) the collection of annual-resolution dendroclimatic, ice core, ice melt, and long historical records used by Bradley and Jones⁶ combined with other coral, ice core, dendroclimatic, and long instrumental records. The long

instrumental records have been formed into annual mean anomalies relative to the 1902–80 reference period, and gridded onto a $5^\circ \times 5^\circ$ grid (yielding 11 temperature grid-point series and 12 precipitation grid-point series dating back to 1820 or earlier) similar to that shown in Fig. 1b. Certain densely sampled regional dendroclimatic data sets have been represented in the network by a smaller number of leading principal components (typically 3–11 depending on the spatial extent and size of the data set). This form of representation ensures a reasonably homogeneous spatial sampling in the multiproxy network (112 indicators back to 1820).

Potential limitations specific to each type of proxy data series must be carefully taken into account in building an appropriate network. Dating errors in a given record (for example, incorrectly assigned annual layers or rings) are particularly detrimental if mutual information is sought to describe climate patterns on a year-by-year basis. Standardization of certain biological proxy records relative to estimated growth trends, and the limits of constituent chronology segment lengths (for example, in dendroclimatic reconstructions), can restrict the maximum timescale of climate variability that is recorded⁹, and only a limited subset of the indicators in the multiproxy network may thus 'anchor in' the longest-term trends (for example, variations on timescales greater than 500 years). However, the dendroclimatic data used were carefully screened for conservative standardization and sizeable segment lengths. Moreover, the mutual information contained in a diverse and widely distributed set of independent climate indicators can more faithfully capture the consistent climate signal that is present, reducing the compromising effects of biases and weaknesses in the individual indicators.

Monthly instrumental land air and sea surface temperature¹⁰ grid-point data (Fig. 1b) from the period 1902–95 are used to calibrate the proxy data set. Although there are notable spatial gaps, this network covers significant enough portions of the globe to form reliable estimates of Northern Hemisphere mean temperature, and certain regional indices of particular importance such as the 'NINO3' eastern tropical Pacific surface temperature index often used to describe the El Niño phenomenon. The NINO3 index is constructed from the eight grid-points available within the conventional NINO3 box (5° S to 5° N, 90° – 150° W).

Multiproxy calibration

Although studies have shown that well chosen regional paleoclimate reconstructions can act as surprisingly representative surrogates for

large-scale climate^{11–13}, multiproxy networks seem to provide the greatest opportunity for large-scale palaeoclimate reconstruction⁶ and climate signal detection¹⁵. There is a rich tradition of multivariate statistical calibration approaches to palaeoclimate reconstruction, particularly in the field of dendroclimatology where the relative strengths and weaknesses of various approaches to multivariate calibration have been well studied^{14,15}. Such approaches have been applied to regional dendroclimatic networks to reconstruct regional patterns of temperature^{16,17} and atmospheric circulation^{18–20} or specific climate phenomena such as the Southern Oscillation²¹. Largely because of the inhomogeneity of the information represented by different types of indicators in a true ‘multiproxy’ network, we found conventional approaches (for example, canonical correlation analysis, CCA, of the proxy and instrumental data sets) to be relatively ineffective. Our approach to climate pattern reconstruction relates closely to statistical approaches which have recently been applied to the problem of filling-in sparse early instrumental climate fields, based on calibration of

the sparse sub-networks against the more widespread patterns of variability that can be resolved in shorter data sets^{22,23}. We first decompose the twentieth-century instrumental data into its dominant patterns of variability, and subsequently calibrate the individual climate proxy indicators against the time histories of these distinct patterns during their mutual interval of overlap. One can think of the instrumental patterns as ‘training’ templates against which we calibrate or ‘train’ the much longer proxy data (that is, the ‘trainee’ data) during the shorter calibration period which they overlap. This calibration allows us to subsequently solve an ‘inverse problem’ whereby best estimates of surface temperature patterns are deduced back in time before the calibration period, from the multiproxy network alone.

Implicit in our approach are at least three fundamental assumptions. (1) The indicators in our multiproxy trainee network are linearly related to one or more of the instrumental training patterns. In the relatively unlikely event that a proxy indicator represents a truly local climate phenomenon which is uncorrelated with larger-

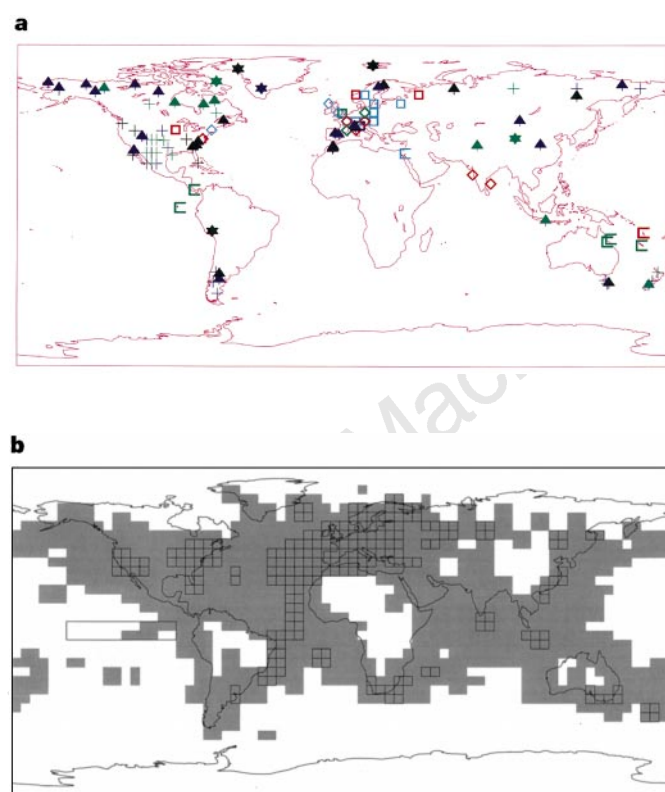


Figure 1 Data used in this study. **a**, Distribution of annual-resolution proxy indicators used in this study. Dendroclimatic reconstructions are indicated by ‘tree’ symbols, ice core/ice melt proxies by ‘star’ symbols and coral records by ‘C’ symbols. Long historical records and instrumental ‘grid-points’ series are shown by squares (temperature) or diamonds (precipitation). Groups of ‘+’ symbols indicate principal components of dense tree-ring sub-networks, with the number of such symbols indicating the number of retained principal components. Sites are shown dating back to at least 1820 (red), 1800 (blue-green), 1750 (green), 1600 (blue) and 1400 (black). Certain sites (for example, the Quelccaya ice core) consist of multiple proxy indicators (for example, multiple cores, and both $\delta^{18}\text{O}$ isotope and accumulation measurements). **b**, Distribution of the 1,082 nearly continuous available land air/sea surface temperature grid-point data available from 1902 onward, indicated by shading. The squares indicate the subset of 219 grid-points with nearly continuous records extending back to 1854 that are used for verification. Northern Hemisphere (NH) and global (GLB) mean temperature are estimated as areally weighted (that is, cosine latitude) averages over the Northern Hemisphere and global domains respectively.

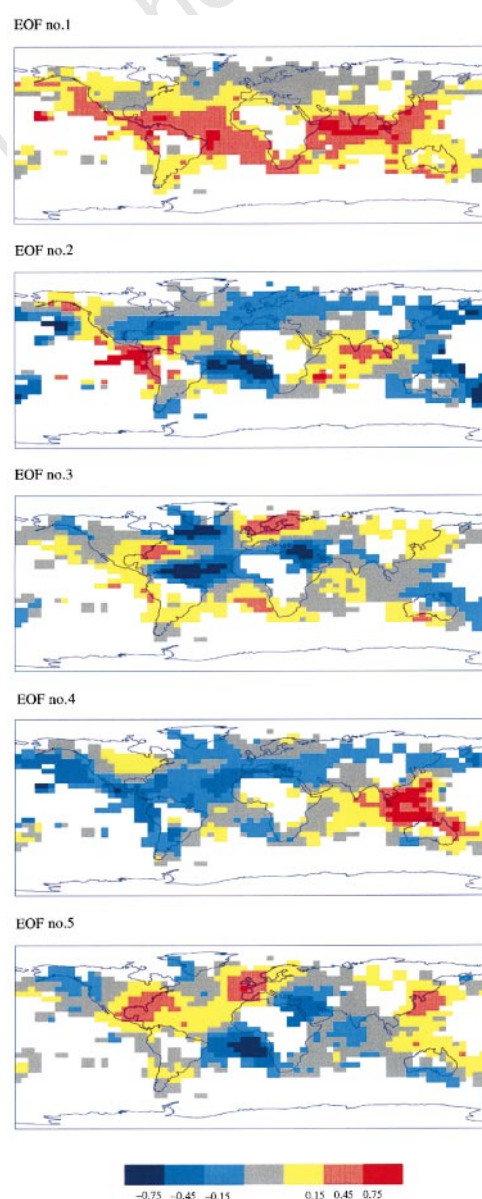


Figure 2 Empirical orthogonal functions (EOFs) for the five leading eigenvectors of the global temperature data from 1902 to 1980. The gridpoint areal weighting factor used in the PCA procedure has been removed from the EOFs so that relative temperature anomalies can be inferred from the patterns.

scale climate variations, or represents a highly nonlinear response to climate variations, this assumption will not be satisfied. (2) A relatively sparse but widely distributed sampling of long proxy and instrumental records may nonetheless sample most of the relatively small number of degrees of freedom in climate patterns at inter-annual and longer timescales. Regions not directly represented in the trainee network may nonetheless be indirectly represented through teleconnections with regions that are. The El Niño/Southern Oscillation (ENSO), for example, shows global-scale patterns of climatic influence²⁴, and is an example of a prominent pattern of variability which, if captured, can potentially describe variability in regions not directly sampled by the trainee data. (3) Patterns of variability captured by the multiproxy network have analogues in the patterns we resolve in the shorter instrumental data. This last assumption represents a fairly weak ‘stationarity’ requirement—we do not require that the climate itself be stationary. In fact, we expect that some sizeable trends in the climate may be resolved by our reconstructions. We do, however, assume that the fundamental spatial patterns of variation which the climate has shown during the past century are similar to those by which it has varied during past recent centuries. Studies of instrumental surface-temperature patterns suggest that such a form of stationarity holds up at least on multidecadal timescales, during the past century²³. The statistical cross-validation exercises we describe later provide the best evidence that these key underlying assumptions hold.

We isolate the dominant patterns of the instrumental surface-temperature data through principal component analysis²⁵ (PCA). PCA provides a natural smoothing of the temperature field in terms of a small number of dominant patterns of variability or ‘empirical eigenvectors’. Each of these eigenvectors is associated with a characteristic spatial pattern or ‘empirical orthogonal function’ (EOF) and its characteristic evolution in time or ‘principal component’ (PC). The ranking of the eigenvectors orders the fraction of variance they describe in the (standardized) multivariate data during the calibration period. The first five of these eigenvectors describe a fraction $\beta = 0.93$ (that is, 93%) of the global-mean (GLB) temperature variations, 85% of the Northern Hemisphere-mean (NH) variations, 67% of the NINO3 index, and 76% of the non-trend-related (DETR) NH variance (see Methods for a description of the β statistic used here as a measure of resolved variance). A sizeable fraction of the total multivariate spatiotemporal variance (MULT) in the raw (instrumental) data (27%) is described by these five eigenvectors, or about 30% of the standardized variance (no. 1 = 12%, no. 2 = 6.5%, no. 3 = 5%, no. 4 = 4%, no. 5 = 3.5%). Figure 2 shows the EOFs of the first five eigenvectors. The associated PCs and their reconstructed counterparts (RPCs) are discussed in the next section. The first eigenvector, associated with the significant global warming trend of the past century, describes much of the variability in the global (GLB = 88%) and hemispheric (NH = 73%) means. Subsequent eigenvectors, in contrast, describe much of the spatial variability relative to the large-scale means (that is, much of the remaining MULT). The second eigenvector is the dominant ENSO-related component, describing 41% of the variance in the NINO3 index. This eigenvector shows a modest negative trend which, in the eastern tropical Pacific, describes a ‘La Niña’-like cooling trend²⁶, which opposes warming in the same region associated with the global warming pattern of the first eigenvector. The third eigenvector is associated largely with interannual-to-decadal scale variability in the Atlantic basin and carries the well-known temperature signature of the North Atlantic Oscillation (NAO)²⁷ and decadal tropical Atlantic dipole²⁸. The fourth eigenvector describes a primarily multidecadal timescale variation with ENSO-scale and tropical/subtropical Atlantic features, while the fifth eigenvector is dominated by multidecadal variability in the entire Atlantic basin and neighbouring regions that has been widely noted elsewhere^{29–34}.

We calibrate each of the indicators in the multiproxy data

network against these empirical eigenvectors at annual-mean resolution during the 1902–80 training interval. Although the seasonality of variability is potentially important—many extratropical proxy indicators, for example, reflect primarily warm-season variability^{6,7}—we seek in the present study to resolve only annual-mean conditions, exploiting the seasonal climate persistence, and the fact that the mutual information from data reflecting various seasonal windows should provide complementary information regarding annual mean climate conditions¹⁰. Following this calibration, we apply an overdetermined optimization procedure to determine the best combination of eigenvectors represented by the multiproxy network back in time on a year-by-year basis, with a spatial coverage dictated only by the spatial extent of the instrumental training data. From the RPCs, spatial patterns and all relevant averages or indices can be readily determined. The details of the entire statistical approach are described in the Methods section.

The skill of the temperature reconstructions (that is, their statistical validity) back in time is established through a variety of complementary independent cross-validation or ‘verification’ exercises (see Methods). We summarize here the main results of these experiments (details of the quantitative results of the calibration and verification procedures are available; see Supplementary Information).

(1) In the reconstructions from 1820 onwards based on the full multiproxy network of 112 indicators, 11 eigenvectors are skilfully resolved (nos 1–5, 7, 9, 11, 14–16) describing ~70–80% of the variance in NH and GLB mean series in both calibration and verification. (Verification is based here on the independent 1854–1901 data set which was withheld; see Methods.) Figure 3 shows the spatial patterns of calibration β , and verification β and the squared correlation statistic r^2 , demonstrating highly significant reconstructive skill over widespread regions of the reconstructed spatial domain. 30% of the full spatiotemporal variance in the gridded data set is captured in calibration, and 22% of the variance is verified in cross-validation. Some of the degradation in the verification score relative to the calibration score may reflect the decrease in instrumental data quality in many regions before the twentieth century rather than a true decrease in resolved variance. These scores thus compare favourably to the 40% total spatiotemporal variance that is described by simply filtering the raw 1902–80 instrumental data with 11 eigenvectors used in calibration, suggesting that the multiproxy calibrations are describing a level of variance in the data reasonably close to the optimal ‘target’ value. Although a verification NINO3 index is not available from 1854 to 1901, correlation of the reconstructed NINO3 index with the available Southern Oscillation index (SOI) data from 1865 to 1901 of $r = -0.38$ ($r^2 = 0.14$) compares reasonably with its target value given by the correlation between the actual instrumental NINO3 and SOI index from 1902 to 1980 ($r = -0.72$). Furthermore, the correspondence between the reconstructed NINO3 index warm events and historical³⁵ El Niño chronology back to 1820 (see Methods) is significant at the 98% level.

(2) The calibrations back to 1760, based on 93 indicators, continue to resolve at least nine eigenvectors (nos 1–5, 7, 9, 11, 15) with no degradation of calibration or verification resolved variance in NH, and only slight degradation in MULT (calibration ~27%, verification ~17%). Our reconstructions are thus largely indistinguishable in skill back to 1760.

(3) The network available back to 1700 of 74 indicators (including only two instrumental or historical indicators) skilfully resolves five eigenvectors (nos 1, 2, 5, 11, 15) and shows some significant signs of decrease in reconstructive skill. In this case, ~60–70% of NH variance is resolved in calibration and verification, ~14–18% of MULT in calibration, and 10–12% of MULT in verification. The verification r of NINO3 with the SOI is in the range of $r \approx -0.25$ to -0.35 , which is statistically significant (as is the correspondence

with the historical³⁵ chronology back to 1700) but notably inferior to the later calibrations. In short, both spatial patterns and large-scale means are skilfully resolved, but with significantly less resolved variance than in later calibrations.

(4) The network of 57 indicators back to 1600 (including one historical record) skilfully resolves four eigenvectors (nos 1, 2, 11, 15). 67% of NH is resolved in calibration, and 53% in verification. 14% of MULT is resolved in calibration, and 12% of MULT in verification. A significant, but modest, level of ENSO-scale variability is resolved in the calibrations.

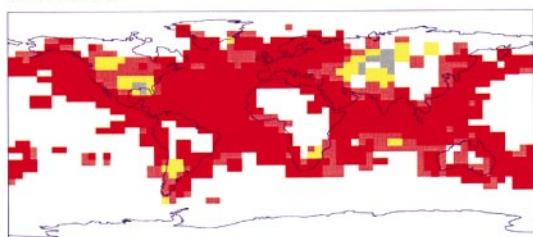
(5) The network of 24 proxy indicators back to 1450 resolves two eigenvectors (nos 1, 2) and ~40–50% of NH in calibration and verification. Only ~10% of MULT is resolved in calibration and

~5% in verification. There is no skilful reconstruction of ENSO-scale variability. Thus spatial reconstructions are of marginal usefulness this far back, though the largest-scale quantities are still skilfully resolved.

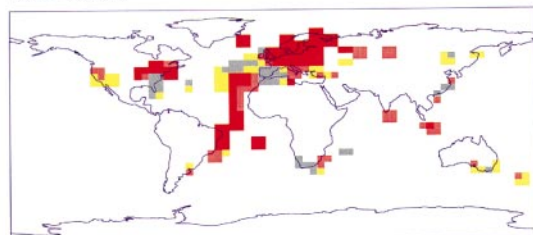
(6) The multiproxy network of 22 indicators available back to 1400 resolves only the first eigenvector, associated with 40–50% of resolved variance in NH in calibration and verification. There is no useful resolution of spatial patterns of variability this far back. The sparser networks available before 1400 show little evidence of skill in reconstructing even the first eigenvector, terminating useful reconstruction at the initial year AD 1400.

(7) Experiments using trainee networks containing only proxy (that is, no instrumental or historical) indicators establish the most

CALIBRATION β



VERIFICATION β



VERIFICATION r^2

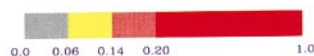
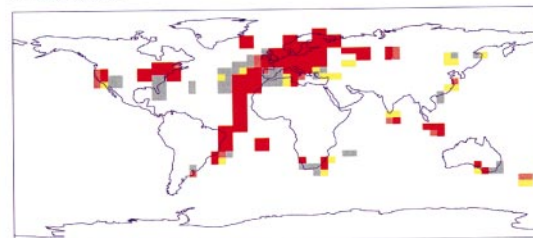
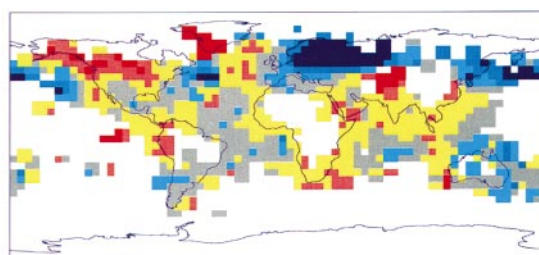
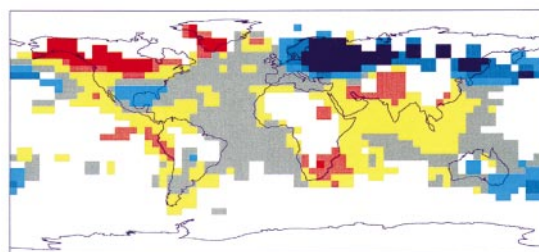


Figure 3 Spatial patterns of reconstruction statistics. Top, calibration β (based on 1902–80 data); middle, verification (based on 1854–1901 data) β ; bottom, verification r^2 (also based on 1854–1901 data). For the β statistic, values that are insignificant at the 99% level are shown in grey negative; but 99% significant values are shown in yellow, and significant positive values are shown in two shades of red. For the r^2 statistic, statistically insignificant values (or any grid-points with unphysical values of correlation $r < 0$) are indicated in grey. The colour scale indicates values significant at the 90% (yellow), 99% (light red) and 99.9% (dark red) levels (these significance levels are slightly higher for the calibration statistics which are based on a longer period of time). A description of significance level estimation is provided in the Methods section.

1941 TEMPERATURE ANOMALY (RAW)



1941 TEMPERATURE ANOMALY (FILTERED)



1941 TEMPERATURE ANOMALY (RECONSTRUCTED)

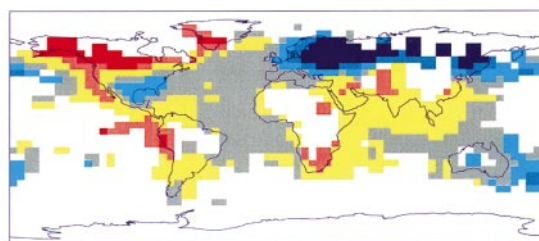


Figure 4 Comparison of the proxy-based spatial reconstructions of the anomaly pattern for 1941 versus the raw data. Comparisons based on actual (top), EOF-filtered (middle), and proxy-reconstructed (bottom) data. Anomalies (relative to 1902–80 climatology) are indicated by the colour scale shown in °C.

truly independent cross-validation of the reconstruction as there is in this case neither spatial nor temporal dependence between the calibration and verification data sets. Such statistically significant verification is demonstrated at the grid-point level (calibration and verification resolved variance $\sim 15\%$ for the MULT statistic), at the largest scales (calibration and verification resolved variance $\sim 60\text{--}65\%$ for NH) and the NINO3-scale (90–95% statistical significance for all verification diagnostics). In contrast, networks containing only the 24 long historical or instrumental records available back to 1820 resolve only $\sim 30\%$ of NH in calibration or verification, and the modest multivariate calibration and verification resolved variance scores of MULT ($\sim 10\%$) are artificially inflated by the high degree of spatial correlation between the instrumental ‘multi-proxy’ predictor and instrumental predictand data. No evidence of skilful ENSO-scale reconstruction is evident in these latter reconstructions. In short, the inclusion of the proxy data in the ‘multi-proxy’ network is essential for the most skilful reconstructions. But certain sub-components of the proxy dataset (for example, the dendroclimatic indicators) appear to be especially important in resolving the large-scale temperature patterns, with notable decreases in the scores reported for the proxy data set if all dendroclimatic indicators are withheld from the multiproxy network. On the other hand, the long-term trend in NH is relatively robust to the inclusion of dendroclimatic indicators in the network, suggesting that potential tree growth trend biases are not influential in the multiproxy climate reconstructions. The network of all combined proxy and long instrumental/historical indicators provide the greatest cross-validated estimates of skilful reconstruction, and are used in obtaining the reconstructions described below.

Temperature reconstructions

The reconstructions discussed here are derived using all indicators available, and using the optimal eigenvector subsets determined in the calibration experiments described above (11 from 1780–1980, 9 from 1760–1779, 8 from 1750–1759, 5 from 1700–1749, 4 from 1600–1699, 2 from 1450–1599, 1 from 1400–1449). To better illustrate the workings and effectiveness of the proxy pattern reconstruction procedure, we show as an example (Fig. 4) the actual, the EOF-filtered, and the reconstructed temperature pat-

terns for a year (1941) during the calibration interval. This year was a known ENSO year, associated with a warm eastern tropical Pacific and a cold central North Pacific. Pronounced cold anomalies were also found over large parts of Eurasia. The proxy-reconstructed pattern captures these features, although in a relatively smoothed sense (describing $\sim 30\%$ of the full variance in that pattern), and is remarkably similar to the raw data once it has been filtered by retaining only the 11 eigenvectors (nos 1–5, 7, 9, 11, 14–16) used in pattern reconstruction. It is thus visually apparent that the multiproxy network is quite capable of resolving much of the structure resolved by the eigenvectors retained in the calibration process.

We consider the temporal variations in the first five RPCs (Fig. 5a). The positive trend in RPC no. 1 during the twentieth century is clearly exceptional in the context of the long-term variability in the associated eigenvector, and indeed describes much of the unprecedented warming trend evident in the NH reconstruction. The negative trend in RPC no. 2 during the past century is also anomalous in the context of the longer-term evolution of the associated eigenvector. The recent negative trend is associated with a pattern of cooling in the eastern tropical Pacific (superimposed on warming associated with the pattern of eigenvector no. 1) which may be a modulating negative feedback on global warming²⁶. RPC no. 5 shows notable multidecadal variability throughout both the modern and pre-calibration interval, associated with the wavelike trend of warming and subsequent cooling of the North Atlantic this century discussed earlier^{29–33} and the longer-term multidecadal oscillations in this region detected in a previous analysis of proxy climate networks⁵. This variability may be associated with ocean–atmosphere processes related to the North Atlantic thermohaline circulation^{4,34}.

The long-term trends in the reconstructed annual mean NH series (Fig. 5b) are quite similar to those of decadal Northern Hemisphere summer temperature reconstructions⁶, showing pronounced cold periods during the mid-seventeenth and nineteenth centuries, and somewhat warmer intervals during the mid-sixteenth and late eighteenth centuries, with almost all years before the twentieth century well below the twentieth-century climatological mean. Taking into account the uncertainties in our NH reconstruction (see Methods), it appears that the years 1990, 1995 and

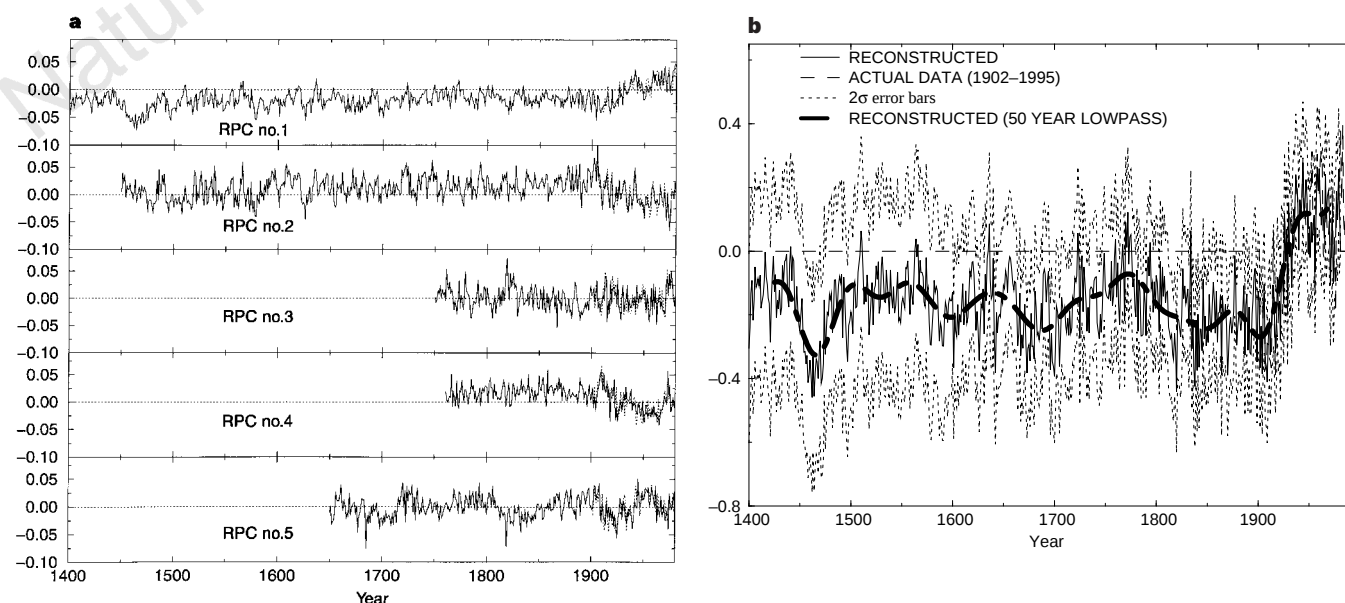


Figure 5 Time reconstructions (solid lines) along with raw data (dashed lines). **a**, For principal components (RPCs) 1–5; **b**, for Northern Hemisphere mean temperature (NH) in $^{\circ}\text{C}$. In both cases, the zero line corresponds to the 1902–80

calibration mean of the quantity. For **b** raw data are shown up to 1995 and positive and negative 2σ uncertainty limits are shown by the light dotted lines surrounding the solid reconstruction, calculated as described in the Methods section.

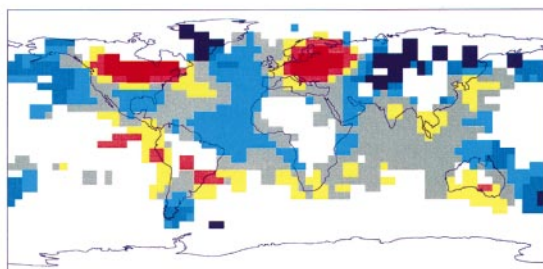
now 1997 (this value recently calculated and not shown) each show anomalies that are greater than any other year back to 1400 at 3 standard errors, or roughly a 99.7% level of certainty. We note that hemispheric mean values are not associated with globally or hemispherically uniform trends. An example of the global pattern for an historically documented³⁵ “very strong” El Niño year (1791) is shown in Fig. 6 top panel, demonstrating the classic warm eastern tropical Pacific and cold central North Pacific sea surface temperature patterns. Analysis of ENSO variability in these reconstructions is discussed in more detail elsewhere³⁶. We also show the reconstructed pattern for 1816 (Fig. 6 bottom panel). Quite anomalous cold is evident throughout much of the Northern Hemisphere (even relative to this generally cold decade) but with a quadrupole pattern of warmth near Newfoundland and the Near East, and enhanced cold in the eastern United States and Europe consistent with the anomalous atmospheric circulation associated with the NAO pattern. Such a pattern is indeed observed in empirical³⁷ and model-based studies³⁸ of the atmospheric response to volcanic forcing. We infer in the 1816 temperature pattern a climatic response to the explosive Tambora eruption of April 1815 based on both the anomalous hemispheric coolness and the superimposed NAO-like pattern. Reconstructed time series RPCs nos 1–5, the NH series, the NINO3 index and reconstructions for specific grid-points

can be obtained through the NOAA palaeoclimatology Web site (<http://www.ngdc.noaa.gov/paleo/paleo.html>).

Attribution of climate forcings

We take an empirical approach to detecting the possible effects of external forcings on the climate. The reconstructed NH series is taken as a diagnostic of the global climate, and we examine its relationship with three candidate external forcings during the period 1610–1995 including (1) CO₂ measurements³⁹ as a proxy for total greenhouse-gas changes, (2) reconstructed solar irradiance variations² and (3) the weighted historical ‘dust veil index’ (DVI) of explosive volcanism (see Fig. 31.1 in ref. 40) updated with recent data⁴¹. While we warn that historical series for these forcing agents are imperfectly known or measured, they do nonetheless represent our best estimates of the time-histories of the corresponding forcings. More detailed discussions of the estimation of, and potential sources of uncertainty or bias in, these series are available^{2,39,40}. Industrial-aerosol forcing of the climate has also been suggested as an important forcing of recent climate^{42,43}, but its physical basis is still controversial⁴⁴, and difficult to estimate observationally. Noting that in any case, this forcing is not believed to be important before about 1940, its omission should be inconsequential in our long-term detection approach. Our empirical

1791 TEMPERATURE ANOMALY PATTERN



1816 TEMPERATURE ANOMALY PATTERN

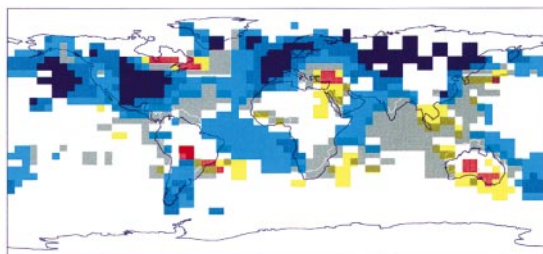


Figure 6 Reconstructed annual temperature patterns for two example years. Top, 1791; bottom, 1816. The colours indicate regions which exceeded (either positively or negatively) the threshold indicated in °C. The zero baseline is defined by the 1902–80 climatological mean for each grid-point.

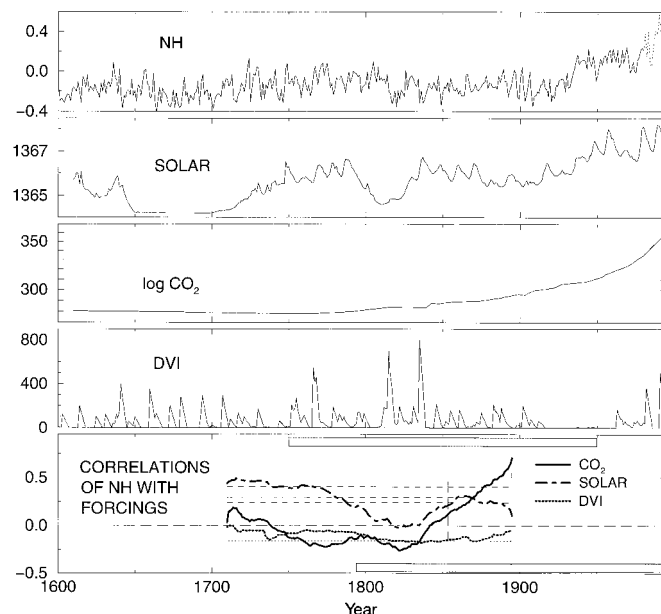


Figure 7 Relationships of Northern Hemisphere mean (NH) temperature with three candidate forcings between 1610 and 1995. Panels, (top to bottom) as follows. ‘NH’, reconstructed NH temperature series from 1610–1980, updated with instrumental data from 1981–95. ‘Solar’, reconstructed solar irradiance. ‘log CO₂’, greenhouse gases represented by atmospheric CO₂ measurements. ‘DVI’, weighted volcanic dust veil index. Bottom panel, evolving multivariate correlation of NH series with the three forcings NH, Solar, log CO₂. The time axis denotes the centre of a 200-year moving window. One-sided (positive) 90%, 95%, 99% significance levels (see text) for correlations with CO₂ and solar irradiance are shown by horizontal dashed lines, while the one-sided (negative) 90% significance threshold for correlations with the DVI series is shown by the horizontal dotted line. The grey bars indicate two difference 200-year windows of data, with the long-dashed vertical lines indicating the centre of the corresponding window.

signal detection is complementary to that of model-based “fingerprint” signal detection studies^{42,43,45}; while our empirical approach relies on the faithfulness of the reconstructed forcing series and on the assumption of a linear and contemporaneous response to forcings, it does not suffer the potential weaknesses of incomplete representations of internal feedback processes²⁶, poorly constrained parametrizations of climatic responses⁴⁴, and underestimated natural variability¹ in model-based studies. To the extent that the response to forcing is not contemporaneous, but rather is delayed owing to the inertia of the slow-response components of the climate system (for example, the ocean and cryosphere), our detection approach will tend to underestimate the response to forcings, making the approach a relatively conservative one.

We estimate the response of the climate to the three forcings based on an evolving multivariate regression method (Fig. 7). This time-dependent correlation approach generalizes on previous studies of (fixed) correlations between long-term Northern Hemisphere temperature records and possible forcing agents^{2,3}. Normalized regression (that is, correlation) coefficients r are simultaneously estimated between each of the three forcing series and the NH series from 1610 to 1995 in a 200-year moving window. The first calculated value centred at 1710 is based on data from 1610 to 1809, and the last value, centred at 1895, is based on data from 1796 to 1995—that is, the most recent 200 years. A window width of 200 yr was chosen to ensure that any given window contains enough samples to provide good signal-to-noise ratios in correlation estimates. Nonetheless, all of the important conclusions drawn below are robust to choosing other reasonable (for example, 100-year) window widths.

We test the significance of the correlation coefficients (r) relative to a null hypothesis of random correlation arising from natural climate variability, taking into account the reduced degrees of freedom in the correlations owing to substantial trends and low-frequency variability in the NH series. The reduced degrees of freedom are modelled in terms of first-order markovian ‘red noise’ correlation structure of the data series, described by the lag-one autocorrelation coefficient ρ during a 200-year window. This parameter ranges from 0.48 in the first window (1610–1809) to 0.77 in the final window (1796–1995) of the moving correlation, the considerably larger recent value associated with the substantial global warming trend of the past century. This latter trend has been shown to be inconsistent with red noise⁴⁶ and could thus itself be argued as indicative of externally forced variability. An argument could in this sense, be made for using the smaller pre-industrial value $\rho = 0.48$ of the NH series in estimating the statistical degrees of freedom appropriate for the null hypothesis of natural variability. Nonetheless, we make the conservative choice of adopting the largest value $\rho = 0.77$ as representative of the natural serial correlation in the series. We use Monte Carlo simulations to estimate the likelihood of chance spurious correlations of such serially correlated noise with each of the three actual forcing series. The associated confidence limits are approximately constant between sliding 200-year windows. For (positive) correlations with both CO₂ and solar irradiance, the confidence levels are both approximately 0.24 (90%), 0.31 (95%), 0.41 (99%), while for the ‘whiter’, relatively trendless, DVI index, the confidence levels for (negative) correlations are somewhat lower (–0.16, –0.20, –0.27 respectively). A one-sided significance test is used in each case because the physical nature of the forcing dictates a unique expected sign to the correlations (positive for CO₂ and solar irradiance variations, negative for the DVI fluctuations).

The correlation statistics indicate highly significant detection of solar irradiance forcing in the NH series during the ‘Maunder Minimum’ of solar activity from the mid-seventeenth to early eighteenth century which corresponds to an especially cold period. In turn, the steady increase in solar irradiance from the early nineteenth century through to the mid-twentieth century

coincides with the general warming over the period, showing peak correlation during the mid-nineteenth century. The regression against solar irradiance indicates a sensitivity to changes in the ‘solar constant’ of $\sim 0.1 \text{ K W}^{-1} \text{ m}^{-2}$, which is consistent with recent model-based studies⁴². Greenhouse forcing, on the other hand, shows no sign of significance until a large positive correlation sharply emerges as the moving window slides into the twentieth century. The partial correlation with CO₂ indeed dominates over that of solar irradiance for the most recent 200-year interval, as increases in temperature and CO₂ simultaneously accelerate through to the end of 1995, while solar irradiance levels off after the mid-twentieth century. It is reasonable to infer that greenhouse-gas forcing is now the dominant external forcing of the climate system. Explosive volcanism exhibits the expected marginally significant negative correlation with temperature during much of 1610–1995 period, most pronounced in the 200-year window centred near 1830 which includes the most explosive volcanic events.

A variety of general circulation^{42,47} and energy-balance model experiments^{43,45,48} as well as statistical comparisons of twentieth-century global temperatures with forcing series⁴⁹ suggest that, although both solar and greenhouse-gas forcings play some role in explaining twentieth-century climate trends, greenhouse gases appear to play an increasingly dominant role during this century. Such a proposition is consistent with the results of this study.

As larger numbers of high-quality proxy reconstructions become available in diverse regions of the globe, it may be possible to assimilate a more globally representative multiproxy data network. Given the high level of skill possible in large-scale reconstruction back to 1400 with the present network, it is reasonable to hope that it may soon be possible to faithfully reconstruct mean global temperatures back over the entire millennium, resolving for example the enigmatic⁷ medieval period. Geothermal measurements from boreholes⁵⁰ recover long-term temperature trends without many of the complications of traditional proxy indicators and, in combination with traditional multiproxy networks, may prove helpful in better resolving trends over many centuries. With a better knowledge of how the climate has varied before the twentieth century, we will be able to place even better constraints on the importance of natural and anthropogenic factors governing the climate of the past few centuries, factors which will no doubt continue to affect climate variability in the future, in addition to any anthropogenic effects. □

Methods

Statistics. We use as our primary diagnostic of calibration and verification reconstructive skill the conventional ‘resolved variance’ statistic;

$$\beta = 1 - \frac{\sum (y_{\text{ref}} - \hat{y})^2}{\sum y_{\text{ref}}^2}$$

where y_{ref} is the reference series (the raw data in the case of calibration or the verification dataset in the case of verification) and $\hat{y}$ is the series being compared to it (the proxy-reconstructed data for either calibration or verification). We compute β for each grid-point, and for the NH, GLB and MULT quantities. The sum extends over the time interval of comparison, and for the multivariate case (MULT), over all gridpoints as well. We also computed a calibration β statistic for the detrended NH series (DETR) to distinguish between explanatory variance associated with the notable trend of the twentieth century, and that related to departures from the trend.

β is a quite rigorous measure of the similarity between two variables, measuring their correspondence not only in terms of the relative departures from mean values (as does the correlation coefficient r) but also in terms of the means and absolute variance of the two series. For comparison, correlation (r) and squared-correlation (r^2) statistics are also determined. The expectation value for two random series is $\beta = -1$. Negative values of β may in fact be statistically significant for sufficient temporal degrees of freedom. Nonetheless, the threshold $\beta = 0$ defines the simple ‘climatological’ model in which a series is assigned its long-term mean. In this sense, statistically significant negative

values of β might still be considered questionable in their predictive or reconstructive skill. Owing to the more rigorous ‘match’ between two series sought by β , highly significant values of β are possible even when r^2 is only marginally significant.

Significance levels were determined for r^2 from standard one-sided tables, accounting for decreased degrees of freedom owing to serial correlation. Significance levels for β were estimated by Monte Carlo simulations, also taking serial correlation into account. Serial correlation is assumed to follow from the null model of AR(1) red noise, and degrees of freedom are estimated based on the lag-one autocorrelation coefficients (ρ) for the two series being compared. Although the values of ρ differ from grid-point to grid-point, this variation is relatively small, making it simplest to use the ensemble average values of ρ over the domain ($\rho \approx 0.2$).

Calibration. With the spatial sampling of $M = 1,082$ continuous monthly grid-point surface temperature anomaly (that is, de-seasonalized) data used (Fig. 1b), the $N = 1,128$ months of data available from 1902 to 1995 were sufficient for a unique, overdetermined eigenvector decomposition (note that $N' = 94$ years of the annual mean data would, in contrast, not be sufficient).

For each grid-point, the mean was removed, and the series was normalized by its standard deviation. A standardized data matrix T of the data is formed by weighting each grid-point by the cosine of its central latitude to ensure areally proportional contributed variance, and a conventional Principal Component Analysis (PCA) is performed,

$$T = \sum_{k=1}^K \lambda_k \mathbf{u}_k \mathbf{v}_k^\dagger$$

decomposing the dataset into its dominant spatiotemporal eigenvectors. The M -vector or empirical orthogonal function (EOF) $\mathbf{v}_k$ describes the relative spatial pattern of the k th eigenvector, the N -vector $\mathbf{u}_k$ or principal component (PC) describes its variation over time, and the scalar λ_k describes the associated fraction of resolved (standardized and weighted) data variance.

In a given calibration exercise, we retain a specified subset of the annually averaged eigenvectors, the annually averaged PCs denoted by $\tilde{\mathbf{u}}_n^k$, where $n = 1, \dots, \tilde{N}$, $\tilde{N} = 79$ is the number of annual averages used of the N -month length data set. In practice, only a small subset N_{cofs} of the highest-rank eigenvectors turn out to be useful in these exercises from the standpoint of verifiable reconstructive skill. An objective criterion was used to determine the particular set of eigenvectors which should be used in the calibration as follows. Preisendorfer’s²⁵ selection rule ‘rule N’ was applied to the multiproxy network to determine the approximate number N_{cofs} of significant independent climate patterns that are resolved by the network, taking into account the spatial correlation within the multiproxy data set. Because the ordering of various eigenvectors in terms of their prominence in the instrumental data, and their prominence as represented by the multiproxy network, need not be the same, we allowed for the selection of non-contiguous sequences of the instrumental eigenvectors. We chose the optimal group of N_{cofs} eigenvectors, from among a larger set (for example, the first 16) of the highest-rank eigenvectors, as the group of eigenvectors which maximized the calibration explained variance. It was encouraging from a consistency standpoint that this subset typically corresponded quite closely to the subset which maximized the verification explained variance statistics (see below), but the objective criterion was, as it should be, independent of the verification process. We emphasize, furthermore, that statistical significance was robustly established, as neither the measures of statistical skill nor the reconstructions themselves were highly sensitive to the precise criterion for selection. In addition to the above means of cross-validation, we also tested the network for sensitivity to the inclusion or elimination of particular trainee data (for example, instrumental/historical records, non-instrumental/historical records, or dendroclimatic proxy indicators).

These N_{cofs} eigenvectors were trained against the N_{proxy} indicators, by finding the least-squares optimal combination of the N_{cofs} PCs represented by each individual proxy indicator during the $\tilde{N} = 79$ year training interval from 1902 to 1980 (the training interval is terminated at 1980 because many of the proxy series terminate at or shortly after 1980). The proxy series and PCs were formed into anomalies relative to the same 1902–80 reference period mean, and the proxy series were also normalized by their standard deviations during that period. This proxy-by-proxy calibration is well posed (that is, a unique optimal solution exists) as long as $\tilde{N} > N_{\text{cofs}}$ (a limit never approached in this study)

and can be expressed as the least-squares solution to the overdetermined matrix equation, $\mathbf{U}\mathbf{x} = \mathbf{y}^{(p)}$, where

$$\mathbf{U} = \begin{bmatrix} \tilde{\mathbf{u}}_1^{(1)} & \tilde{\mathbf{u}}_1^{(2)} & \dots & \tilde{\mathbf{u}}_1^{(N_{\text{cofs}})} \\ \tilde{\mathbf{u}}_2^{(1)} & \tilde{\mathbf{u}}_2^{(2)} & \dots & \tilde{\mathbf{u}}_2^{(N_{\text{cofs}})} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{u}}_{\tilde{N}}^{(1)} & \tilde{\mathbf{u}}_{\tilde{N}}^{(2)} & \dots & \tilde{\mathbf{u}}_{\tilde{N}}^{(N_{\text{cofs}})} \end{bmatrix}$$

is the matrix of annual PCs, and

$$\mathbf{y}^{(p)} = \begin{bmatrix} y^{(p)}_1 \\ y^{(p)}_2 \\ \vdots \\ y^{(p)}_{\tilde{N}} \end{bmatrix}$$

is the time series $\tilde{N}$ -vector for proxy record p .

The N_{cofs} -length solution vector $\mathbf{x} = \mathbf{G}^{(p)}$ is obtained by solving the above overdetermined optimization problem by singular value decomposition for each proxy record $p = 1, \dots, P$. This yields a matrix of coefficients relating the different proxies to their closest linear combination of the N_{cofs} PCs;

$$\mathbf{G} = \begin{bmatrix} G_1^{(1)} & G_2^{(1)} & \dots & G_{N_{\text{cofs}}}^{(1)} \\ G_1^{(2)} & G_2^{(2)} & \dots & G_{N_{\text{cofs}}}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ G_1^{(P)} & G_2^{(P)} & \dots & G_{N_{\text{cofs}}}^{(P)} \end{bmatrix}$$

This set of coefficients will not provide a single consistent solution, but rather represents an overdetermined relationship between the optimal weights on each of the N_{cofs} PCs and the multiproxy network.

Proxy-reconstructed patterns are thus obtained during the pre-calibration interval by the year-by-year solution of the overdetermined matrix equation, $\mathbf{G}\mathbf{z} = \mathbf{y}_{(j)}$, where $\mathbf{y}_{(j)}$ is the predictor vector of values of each of the P proxy indicators during year j . The predictand solution vector $\mathbf{z} = \hat{\mathbf{U}}$ contains the least-squares optimal values of each of the N_{cofs} PCs for a given year. This optimization is overdetermined (and thus well constrained) as long as $P > N_{\text{cofs}}$ which is always realized in this study. It is noteworthy that, unlike conventional palaeoclimate transfer function approaches, there is no specific relationship between a given proxy indicator and a given predictand (that is, reconstructed PC). Instead, the best common choice of values for the small number of N_{cofs} predictands is determined from the mutual information present in the multiproxy network during any given year. The reconstruction approach is thus relatively resistant to errors or biases specific to any small number of indicators during a given year.

This yearly reconstruction process leads to annual sequences of the optimal reconstructions of the retained PCs, which we term the reconstructed principal components or RPCs and denote by $\hat{\mathbf{u}}^k$. Once the RPCs are determined, the associated temperature patterns are readily obtained through the appropriate eigenvector expansion,

$$\hat{T} = \sum_{k=1}^{N_{\text{cofs}}} \lambda_k \hat{\mathbf{u}}^k \mathbf{v}_k$$

while quantities of interest (for example, NH) are calculated from the appropriate spatial averages, and appropriate calibration and verification resolved variance statistics are calculated from the raw and reconstructed data.

Several checks were performed to ensure a reasonably unbiased calibration procedure. The histograms of calibration residuals were examined for possible heteroscedasticity, but were found to pass a χ^2 test for gaussian characteristics at reasonably high levels of significance (NH, 95% level; NINO3, 99% level). The spectra of the calibration residuals for these quantities were, furthermore, found to be approximately ‘white’, showing little evidence for preferred or deficiently resolved timescales in the calibration process. Having established reasonably unbiased calibration residuals, we were able to calculate uncertainties in the reconstructions by assuming that the unresolved variance is gaussian distributed over time. This variance increases back in time (the increasingly sparse multiproxy network calibrates smaller fractions of variance), yielding error bars which expand back in time.

Verification. Verification resolved variance statistics (β) were determined based on two distinct verification data sets including (1) the sparse subset of the

gridded data ($M' = 219$ grid-points) for which independent values are available from 1854 to 1901 (see Fig. 1b) and (2) the small subset of 11 very long instrumental estimated temperature grid-point averages (10 in Eurasia, 1 in North America—see Fig. 1a) constructed from the longest available station measurements. Each of these 'grid-point' series shared at least 70% of their variance with the corresponding temperature grid-point available from 1854–1980, providing verification back to at least 1820 in all cases (and back through the mid and early eighteenth century in many cases). Note that this latter verification data set is only temporally, but not spatially, independent of the multiproxy network itself, which contains these long instrumental grid-point series as a small subset of the network. In case (1), NH and GLB verification statistics are computed as well as the multivariate (MULT) grid-point level verification statistic, although these quantities represent different spatial samplings from those in the full calibration data set owing to the sparser sampling of the verification period. Case (2) provides a longer-term, albeit an even less spatially representative, multivariate verification statistic (MULTb). In this case, the spatial sampling does not permit meaningful estimates of NH or GLB mean quantities. In any of these diagnostics, a positive value of β is statistically significant at >99% confidence as established from Monte Carlo simulations. Verification skills for the NINO3 reconstructions are estimated by other means, as the actual NINO3 index is not available far beyond the beginning of the calibration period. The (negative) correlation r of NINO3 with the SOI annual-mean from 1865 to 1901 (P. D. Jones, personal communication), and a squared congruence statistic g^2 measuring the categorical match between the distribution of warm NINO3 events and the distribution of warm episodes according to the historical³⁵ chronology (available back to the beginning of 1525), were used for statistical cross-validation based on one-sided tables and Monte Carlo simulations, respectively. The results of all calibration and verification experiments are available; see Supplementary Information.

Received 9 May 1997; accepted 27 February 1998.

- Barnett, T. P., Santer, B., Jones, P. D., Bradley, R. S. & Briffa, K. R. Estimates of low frequency natural variability in near-surface air temperature. *Holocene* **6**, 255–263 (1996).
- Lean, J., Beer, J. & Bradley, R. S. Reconstruction of solar irradiance since 1610: Implications for climate change. *Geophys. Res. Lett.* **22**, 3195–3198 (1995).
- Crowley, T. J. & Kim, K. Y. Comparison of proxy records of climate change and solar forcing. *Geophys. Res. Lett.* **23**, 359–362 (1996).
- Delworth, T. D., Manabe, S. & Stouffer, R. J. Interdecadal variations of the thermohaline circulation in a coupled ocean–atmosphere model. *J. Clim.* **6**, 1993–2011 (1993).
- Mann, M. E., Park, J. & Bradley, R. S. Global interdecadal and century-scale oscillations during the past five centuries. *Nature* **378**, 266–270 (1995).
- Bradley, R. S. & Jones, P. D. 'Little Ice Age' summer temperature variations: their nature and relevance to recent global warming trends. *Holocene* **3**, 367–376 (1993).
- Hughes, M. K. & Diaz, H. F. Was there a 'Medieval Warm Period' and if so, where and when? *Clim. Change* **26**, 109–142 (1994).
- Diaz, H. F. & Pulwarty, R. S. An analysis of the time scales of variability in centuries-long ENSO-sensitive records in the last 1000 years. *Clim. Change* **26**, 317–342 (1994).
- Cook, E. R., Briffa, K. R., Meko, D. M., Graybill, D. A. & Funkhouser, G. The 'segment length' curse in long tree-ring chronology development for paleoclimatic studies. *Holocene* **5**, 229–237 (1995).
- Jones, P. D. & Briffa, K. R. Global surface air temperature variations during the 20th century: Part 1—Spatial, temporal and seasonal details. *Holocene* **1**, 165–179 (1992).
- Bradley, R. S. in *Climatic Variations and Forcing Mechanisms of the Last 2000 Years* (eds Jones, P. D., Bradley, R. S. & Jouzel, J.) 603–624 (Springer, Berlin, 1996).
- Jacoby, G. C. & D'Arrigo, R. Reconstructed Northern Hemisphere annual temperature since 1671 based on high-latitude tree-ring data from North America. *Clim. Change* **14**, 39–59 (1989).
- Jacoby, G. C., D'Arrigo, R. D. & Tsevegyn, D. Mongolian tree rings and 20th-century warming. *Science* **9**, 771–773 (1996).
- Fritts, H. C., Blasing, T. J., Hayden, B. P. & Kutzbach, J. E. Multivariate techniques for specifying tree-growth and climate relationships and for reconstructing anomalies in paleoclimate. *J. Appl. Meteorol.* **10**, 845–864 (1971).
- Cook, E. R., Briffa, K. R. & Jones, P. D. Spatial regression methods in dendroclimatology: a review and comparison of two techniques. *Int. J. Climatol.* **14**, 379–402 (1994).
- Briffa, K. R., Jones, P. D. & Schweingruber, F. H. Tree-ring density reconstructions of summer temperature pattern across western North America since 1600. *J. Clim.* **5**, 735–753 (1992).
- Schweingruber, F. H., Briffa, K. R. & Jones, P. D. Yearly maps of summer temperatures in western Europe from A.D. 1750 to 1975 and Western North American from 1600 to 1982. *Vegetatio* **92**, 5–71 (1991).
- Fritts, H. *Reconstructing Large-scale Climatic Patterns from Tree Ring Data* (Univ. Arizona Press, Tucson, 1991).

- Guiot, J. The combination of historical documents and biological data in the reconstruction of climate variations in space and time. *Palaeoklimaforschung* **7**, 93–104 (1988).
- D'Arrigo, R. D., Cook, E. R., Jacoby, G. C. & Briffa, K. R. NAO and sea surface temperature signatures in tree-ring records from the North Atlantic sector. *Quart. Sci. Rev.* **12**, 431–440 (1993).
- Stahle, D. W. & Cleaveland, M. K. Southern Oscillation extremes reconstructed from tree rings of the Sierra Madre Occidental and southern Great Plains. *J. Clim.* **6**, 129–140 (1993).
- Smith, T. M., Reynolds, R. W., Livezey, R. E. & Stokes, D. C. Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Clim.* **9**, 1403–1420 (1996).
- Kaplan, A. *et al.* Analyses of global sea surface temperature 1856–1991. *J. Geophys. Res.* (in the press).
- Halpert, M. S. & Ropelewski, C. F. Surface temperature patterns associated with the Southern Oscillation. *J. Clim.* **5**, 577–593 (1992).
- Preisendorfer, R. W. *Principal Component Analysis in Meteorology and Oceanography* (Elsevier, Amsterdam, 1988).
- Cane, M. *et al.* Twentieth-century sea surface temperature trends. *Science* **275**, 957–960 (1997).
- Hurrell, J. W. Decadal trends in the North Atlantic Oscillation, regional temperatures and precipitation. *Science* **269**, 676–679 (1995).
- Chang, P., Ji, L. & Li, H. A decadal climate variation in the tropical Atlantic Ocean from thermodynamic air–sea interactions. *Nature* **385**, 516–518 (1997).
- Folland, C. K., Parker, D. E. & Kates, F. E. Worldwide marine temperature fluctuations 1856–1981. *Nature* **310**, 670–673 (1984).
- Kushnir, Y. Interdecadal variations in North Atlantic sea surface temperature and associated atmospheric conditions. *J. Clim.* **7**, 141–157 (1994).
- Schlesinger, M. E. & Ramankutty, N. An oscillation in the global climate system of period 65–70 years. *Nature* **367**, 723–726 (1994).
- Mann, M. E. & Park, J. Global-scale modes of surface temperature variability on interannual to century timescales. *J. Geophys. Res.* **99**, 25819–25833 (1994).
- Mann, M. E. & Park, J. Joint spatiotemporal modes of surface temperature and sea level pressure variability in the northern hemisphere during the last century. *J. Clim.* **9**, 2137–2162 (1996).
- Delworth, T. D., Manabe, S. & Stouffer, R. J. Multidecadal climate variability in the Greenland Sea and surrounding regions: a coupled model simulation. *Geophys. Res. Lett.* **24**, 257–260 (1997).
- Quinn, W. H. & Neal, V. T. in *Climate Since A.D. 1500* (eds Bradley, R. S. & Jones, P. D.) 623–648 (Routledge, London, 1992).
- Mann, M. E., Bradley, R. S. & Hughes, M. K. in *El Niño and the Southern Oscillation: Multiscale Variability and its Impacts on Natural Ecosystems and Society* (eds Diaz, H. F. & Markgraf, V.) (Cambridge Univ. Press, in the press).
- Kelly, P. M., Jones, P. D. & Penggun, J. The spatial response of the climate system to explosive volcanic eruptions. *Int. J. Climatol.* **16**, 537–550 (1996).
- Kirchner, I. & Graf, H. F. Volcanoes and El Niño: signal separation in Northern Hemisphere winter. *Clim. Dyn.* **11**, 341–358 (1995).
- Raynaud, D., Barnola, J. M., Chappellaz, J. & Martinerie, P. in *Climatic Variations and Forcing Mechanisms of the Last 2000 Years* (eds Jones, P. D., Bradley, R. S. & Jouzel, J.) 547–562 (Springer, Berlin, 1996).
- Bradley, R. S. & Jones, P. D. in *Climate Since A.D. 1500* (eds Bradley, R. S. & Jones, P. D.) 606–622 (Routledge, London, 1992).
- Robock, A. & Free, M. P. Ice cores as an index of global volcanism from 1850 to the present. *J. Geophys. Res.* **100**, 11549–11567 (1995).
- Hegerl, G. C. *et al.* Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change. *Clim. Dyn.* **13**, 613–634 (1997).
- North, G. R. & Stevens, M. J. Detecting climate signals in the surface temperature record. *J. Clim.* (in the press).
- Hansen, J., Sato, M. & Ruedy, R. The missing climate forcing. *Phil. Trans. R. Soc. Lond. B* **352**, 231–240 (1997).
- Stevens, M. J. & North, G. R. Detection of the climate response to the solar cycle. *J. Clim.* (in the press).
- Mann, M. E. & Lees, J. Robust estimation of background noise and signal detection in climatic time series. *Clim. Change* **33**, 409–445 (1996).
- Rind, D. & Overpeck, J. Hypothesized causes of decadal-to-century climate variability: climate model results. *Quart. Sci. Rev.* **12**, 357–374 (1993).
- Wigley, T. M. L. & Raper, S. C. B. Natural variability of the climate system and detection of the greenhouse effect. *Nature* **344**, 324–327 (1990).
- Thomson, D. J. Dependence of global temperatures on atmospheric CO₂ and solar irradiance. *Proc. Natl Acad. Sci. USA* **94**, 8370–8377 (1997).
- Pollack, H. N. & Chapman, D. S. Underground records of changing climate. *Sci. Am.* **268**, 44–50 (1993).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

Acknowledgements. This work benefited from discussions with M. Cane, E. Cook, M. Evans, A. Kaplan and collaborators at the Lamont Doherty Earth Observatory. We acknowledge discussions with K. Briffa, T. Crowley, P. Jones, S. Manabe, R. Saravanan and K. Trenberth, as well as the comments of G. Hegerl. We thank R. D'Arrigo, D. Fisher, G. Jacoby, J. Lean, A. Robock, D. Stahle, C. Stockton, E. Vaganov, R. Villalba and the numerous contributors to the International Tree-Ring Data Bank and other palaeoclimate researchers who have made their data available to us for use in this study; we also thank F. Keimig, M. Munro, R. Holmes and C. Aramann for their technical assistance. This work was supported by the NSF and the US Department of Energy. M.E.M. acknowledges support through the Alexander Hollaender Distinguished Postdoctoral Research Fellowship program of the Department of Energy. This work is a contribution to the NSF- and NOAA-sponsored Analysis of Rapid and Recent Climatic Change (ARRCC) project.

Correspondence and requests for materials should be addressed to M.E.M. (e-mail: mann@snow.geo.umass.edu).

Submillimetre images of dusty debris around nearby stars

Wayne S. Holland*, Jane S. Greaves*, B. Zuckerman†, R. A. Webb‡, Chris McCarthy†, Iain M. Coulson*, D. M. Walther‡, William R. F. Dent§, Walter K. Gear|| & Ian Robson*

* Joint Astronomy Centre, 660 N. A'ohōkū Place, Hilo, Hawaii 96720, USA

† Department of Physics & Astronomy, University of California, Los Angeles, Los Angeles, California 90095, USA

‡ Gemini 8m Telescopes Project, 180 N. Kinoole Street No. 207, Hilo, Hawaii 96720, USA

§ Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

|| Mullard Space Science Laboratory, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK

Indirect detections of massive—presumably Jupiter-like—planets orbiting nearby Sun-like stars have recently been reported^{1,2}. Rocky, Earth-like planets are much more difficult to detect, but clues to their possible existence can nevertheless be obtained from observations of the circumstellar debris disks of dust from which they form. The presence of such disks has been inferred³ from excess far-infrared emission but, with the exception of β Pictoris⁴, it has proved difficult to image these structures directly as starlight dominates the faint light scattered by the dust⁵. A more promising approach is to attempt to image the thermal emission from the dust grains at submillimetre wavelengths^{6,7}. Here we present images of such emission around Fomalhaut, β Pictoris and Vega. For each star, dust emission is detected from regions comparable in size to the Sun's Kuiper belt of comets. The total dust mass surrounding each star is only a few lunar masses, so any Earth-like planets present must already have formed. The presence of the central cavity, approximately the size of Neptune's orbit, that we detect in the emission from Fomalhaut may indeed be the signature of such planets.

The observations were made during the period April–October 1997, using the Submillimetre Common-User Bolometer Array (SCUBA)⁸ at the James Clerk Maxwell Telescope (JCMT) on Mauna Kea, Hawaii. Although data were taken at wavelengths of 450 and 850 μm simultaneously, only the 850- μm data are discussed here, as observing conditions were poor for good-quality 450- μm work. The full-width at half-maximum (FWHM) beam size at 850 μm was 14 arcsec, and the array field of view has a diameter of 2.3 arcmin. Telescope pointing was checked before and after each observation on a nearby quasar, with r.m.s. pointing errors of <2 arcsec. The data were analysed using the SCUBA User Reduction Facility⁹. Table 1 details the integration times and flux measurements, and the final maps are presented in Fig. 1. The fluxes

quoted in Table 1 are from the raw maps, whilst the images shown in Fig. 1 have been smoothed slightly by a 7-arcsec gaussian (FWHM) to improve the signal-to-noise ratio.

The 850- μm image of Fomalhaut (Fig. 1a) shows that the dust emission peaks at two positions offset from the star by ~ 10 arcsec to the north and the south. This image is consistent with an edge-on torus ('doughnut') structure with a central cavity with much less dust emission. At the outer edge it is not possible to distinguish between an abrupt drop or a moderate density decline with radius. It was anticipated that the dusty disk would be seen nearly edge-on because Fomalhaut's projected rotation velocity of 100 km s^{-1} (ref. 10) is consistent with a spin axis nearly in the plane of the sky. After deconvolution from the 14-arcsec beam (at the map half-power points), the major axis is 41 ± 2 arcsec, or 315 ± 15 astronomical units (AU; where 1 AU is the mean Sun–Earth distance) at a position angle (PA) of $162 \pm 4^\circ$. The minor axis is $\sim 18 \pm 2$ arcsec deconvolved, and hence implies a torus inclination angle of $64 \pm 5^\circ$ (an inclination angle of 0° indicates a spin axis pointing along our line of sight). These dimensions compare favourably with previous estimates¹¹ at 100 μm . The data also agree well with the low signal-to-noise point-by-point map and flux estimates of an earlier study at the JCMT⁶.

The regions of peak dust emission are ~ 80 AU from the star. This is comparable to the distance of the Kuiper belt of comets in our own Solar System which begins just beyond the orbit of Pluto¹². However, the dust grains are warmer than in the Kuiper belt because Fomalhaut is ~ 16 times more luminous than the Sun. At a distance of 80 AU from Fomalhaut, particles sufficiently large (diameter $\sim 60 \mu\text{m}$) to radiate like black-bodies at wavelengths of 60–100 μm will be at ~ 70 K. (In the Solar System this value is measured at Uranus' distance from the Sun.) This is the temperature that characterizes the far-infrared emission from Fomalhaut¹⁰, so the far-infrared and submillimetre data are consistent with a grain population that is well mixed in space. A variety of grain sizes is needed within this population to explain the spectral energy distribution⁶.

The central cavity is ~ 8 arcsec (60 AU) in diameter, approximately the size of Neptune's orbit. This observed size agrees reasonably well with the size of the inner cavity deduced from models of the far-infrared emission¹³. One plausible explanation is that this region has been cleared of gas and dust by the formation of planetesimals. However, other mechanisms could produce a central cavity; for example, sublimation of ice-mantle grains¹⁴, or radiation-grain drag (Poynting–Robertson effect)⁶. We also note that as cratering records in the inner Solar System indicate debris clearing times significantly longer than the estimated age of Fomalhaut (200 Myr; ref. 15), planets may exist even within the dusty region. If so, their gravity could extend the vertical distribution of particles, enabling the dust to absorb more stellar light and thus enhancing the far-infrared emission¹⁶.

β Pic is one of the youngest stars in the solar vicinity at an age of ~ 10 –100 Myr (ref. 17). The 850- μm map (Fig. 1b) shows a clear

Table 1 Source and observation summary

Star	RA dec. (J2000)	Type	Distance* (pc)	Integration time (h)	Dust peak† δRA $\delta\text{dec.}$ (arcsec)		850- μm flux (mJy per beam)	Dust mass‡ (M_{moon})
α Piscis Austrinus (Fomalhaut)	22 h 57 min 39.0 s –29° 37' 20.0"	A3V	7.69	5.0	3.5W 3.0E Integrated§:	9.7N 9.0S	28.9 $\pm$ 3.7 28.0 $\pm$ 3.7 81.0 $\pm$ 7.2	1.5
β Pictoris	05 h 47 min 17.1 s –51° 04' 00.0"	A5V	19.3	5.3	0.8N 21.4W Integrated§:	9.7N 25.9S	58.3 $\pm$ 6.5 19.1 $\pm$ 6.5 104.3 $\pm$ 10.0	7.8
α Lyrae (Vega)	18 h 36 min 56.3 s +38° 47' 01.2"	A0V	7.76	8.6	5.9E Integrated§:	7.1N	17.3 $\pm$ 3.0 45.7 $\pm$ 5.4	0.7

* Distances are from the Hipparchos catalogue (ESA SP-1200, 1997).

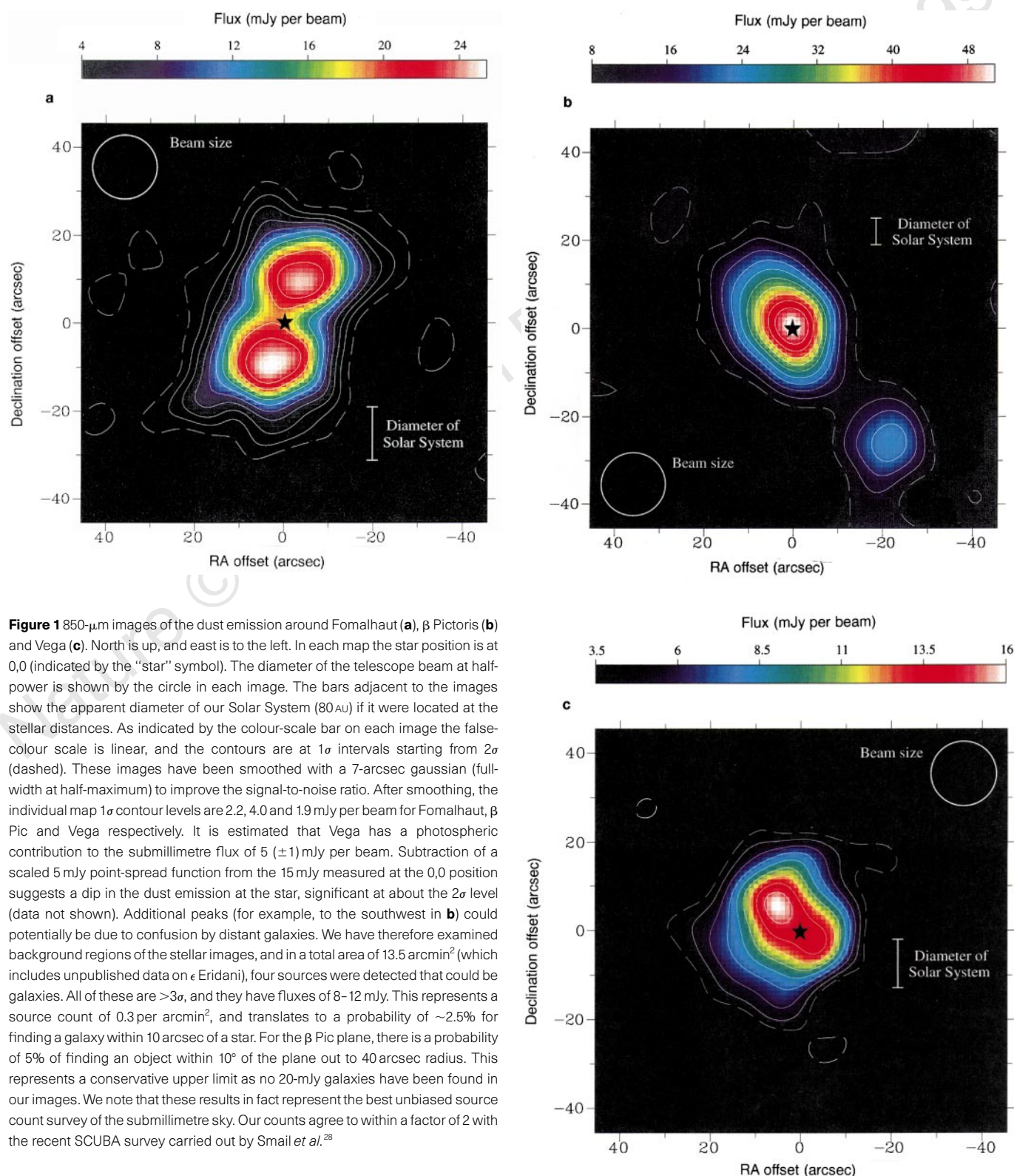
† Estimated errors in the dust peak centroid positions are ± 1 arcsec.

‡ Estimated dust masses are given in lunar masses ($M_{\text{moon}} = 7 \times 10^{25}$ kg).

§ Integrated fluxes (mJy) are measured within a radius of 30 arcsec (40 arcsec for β Pictoris) from the star.

elongation of the dust emission with a PA of $32 \pm 4^\circ$, which, given the resolution of the JCMT beam, is consistent with optical and near-infrared images (for example, $30.7 \pm 0.7^\circ$; ref. 18). The dust emission peaks on the stellar position and extends along the major axis to an observed radius of ~ 13 arcsec (250 AU). The deconvolved size of the disk is $22 \times 11 (\pm 3)$ arcsec (that is, unresolved in the minor axis), giving an inclination angle of $>60^\circ$.

In addition to the main disk, separated patches of emission are seen further out in Fig. 1b, two of which are fairly closely aligned with the plane of the disk. The most prominent of these (RA 21 W, dec 26 S) lies at a PA of $37 (\pm 6)^\circ$ and has a flux of 19.1 mJy per beam. This is almost certainly a real detection and has a PA consistent with a feature in the disk plane. This could be a fragmented outer part of the disk (although such a feature does not appear in optical images),



or possibly a second disk about a low-mass companion (which would explain why there is no peak of corresponding brightness to the northeast). If the age of the companion is <10 Myr, then such a disk might not have dissipated, and could be primordial like those around T-Tauri stars, in contrast to the much older debris disks seen around Fomalhaut and Vega. However, recent deep R-band data at a wavelength of 680 nm revealed no companion to a limiting magnitude of ~ 22 (compare with an R-band magnitude ~ 4 for β Pic) within 6 arcsec of the southwest offset source (P. Kalas and B. A. Smith, personal communications). On the basis of the cooling models of Burrows *et al.*²⁴, and an assumed age of 30 Myr for β Pic, the mass of any companion at the offset source position is less than 10 Jupiter masses. Finally, we cannot rule out a coincidental alignment with a background object, such as a distant galaxy, and this possibility is further discussed in the legend to Fig. 1.

The 850- μ m map of Vega (Fig. 1c) shows an extended, approximately circular structure with dimensions $24 \times 21 (\pm 3)$ arcsec (deconvolved from the beam size), somewhat smaller than the $35 (\pm 5)$ arcsec derived from far-infrared measurements¹⁹. Within this is an elongated bright central region orientated northeast–southwest. The peak is 9 arcsec (70 AU) to the northeast of the star's position, and this is unlikely to be an artefact, as it is consistent with the low signal-to-noise detection in the earlier JCMT study⁶. This region has a superficial resemblance to the Fomalhaut image. However, interpreting this as a torus or edge-on disk encounters two serious problems. The low projected rotational velocity (22 km s^{-1}) and a spectral line analysis²⁰ suggest that the star is rapidly rotating and viewed close to pole-on. Also, the extended dust emission is roughly circular, favouring a pole-on geometry.

The northeast peak is brighter than the corresponding emission southwest of the star by $\sim 2\sigma$, and so it is unclear whether these are separate objects or part of a symmetrical structure. A symmetrical elongation could be due to perturbation of a disk by a planet, as hypothesized for the warped disk of β Pic²¹. Alternatively, a single peak could be interpreted as an object in orbit about Vega,

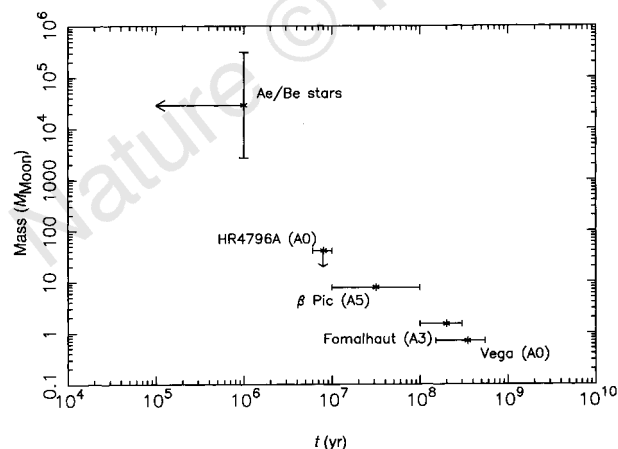


Figure 2 Circumstellar dust masses plotted as a function of stellar age (t). The data for Fomalhaut, β Pic and Vega (stellar types A0 to A5) are compared to JCMT observations of younger high-mass stars. These include an upper limit for HR4796A²⁹ and detections of a sample of emission-line A and B type stars³⁰. These stars have ages of only 0.1–1 Myr, and are often still associated with their parental molecular clouds. Horizontal error bars indicate the uncertainties in the stellar ages, and for the Ae/Be stars, the vertical error bar is the range of masses around several stars. Mass uncertainties depend on the assumed grain temperatures and absorption coefficients. The plotted values could be up to a factor of 2 lower (if all the grains are at the highest dust temperatures measured in the far-infrared¹⁰), and upper limits are essentially unconstrained, as very large grains could contribute substantial mass, while adding little to the total emitting area at submillimetre wavelengths.

surrounded by a concentration of dust. However, circumplanetary disks (such as the one from which Jupiter's moons formed) are expected to dissipate much more rapidly than circumstellar disks, making the presence of such a structure unlikely at Vega's age of 350 Myr (ref. 22). Alternatively, if rocky planetary bodies have formed, the collision of two such objects could inject new dust into the disk, but the objects involved would have to be exceedingly large 'planetesimals' to shed a substantial fraction of a lunar mass in dust. Additional observations were made to investigate this offset peak: first, it was detected by SCUBA at 1,350 μ m with single point photometry. The 850- and 1,350- μ m fluxes imply a spectral index of 2.7, consistent with previous observations of dust around young stars²³. Second, a coronagraphic image of Vega at 2.2 μ m, using the Near-Infrared Camera on the Keck 10-m telescope, revealed no companion down to a limiting magnitude of 16. This rules out orbiting companions with masses greater than 12 Jupiter masses, based on models of 350-Myr-old 'superplanets'²⁴.

We have compared the amounts of dust observed both to the Solar System and to other stars with younger ages. The ages of Fomalhaut, β Pic and Vega—tens to hundreds of millions of years—span a range of particular interest in the history of our Solar System. The formation of giant planets probably occurred within ~ 10 Myr, and the Earth within 100 Myr. In addition, the period of heavy bombardment in the inner Solar System by cometary and asteroidal-sized objects lasted ~ 600 Myr (ref. 25).

Dust masses were estimated (Table 1) from the 850- μ m integrated fluxes using grain temperatures deduced from far-infrared data¹⁰. As the composition and size of the circumstellar grains remain uncertain, we adopt an 850- μ m mass absorption coefficient for dust (K_{abs}) of $1.7 \text{ cm}^2 \text{ g}^{-1}$ for consistency and ease of comparison with the previous best measured dust masses⁶ (the range of values of K_{abs} is further discussed in ref. 26). There may be an associated gas component, but, for example, a sensitive search for atomic hydrogen around β Pic has shown that the gas-to-dust ratio is at least 10 times lower than in the interstellar medium, suggesting most of the gas has been dispersed²⁷. The derived dust masses are of the order of the Moon's mass, and agree well with the previous values⁶. These masses are plotted in Fig. 2 as a function of stellar age. The apparent decrease in the mass of dust particles with time can be attributed to accumulation into planets, or dispersal by mechanisms such as the Poynting–Robertson effect, stellar winds, and sublimation. Whichever of these mechanisms dominates, the dust masses appear to decrease rapidly between ages of 1 and 100 Myr. Figure 2 suggests that circumstellar dust falls below an Earth mass (80 Moon masses) at an estimated age of 10 Myr. Thus at the ages of Fomalhaut, β Pic and Vega, planetesimal accumulation must be well underway if planetary systems are to form.

Our submillimetre images have provided the first detailed information on the morphology of the dust emission around nearby stars. Ice sublimation and Poynting–Robertson drag need to be investigated as possible causes of the central cavity at Fomalhaut, before planet formation becomes the adopted model. If the offset peaks for Vega and β Pic are physically associated with these stars, then compelling models need to be devised. We plan further submillimetre imaging at 450 μ m, with 8-arcsec resolution. Such data may reveal details such as warping, a possible signature of planetary perturbations, and, when combined with longer-wavelength results, should enable investigations of grain growth processes. □

Received 9 January; accepted 17 February 1998.

1. Mayor, M. & Queloz, D. A. Jupiter-sized companion to a solar-type star. *Nature* **378**, 355–359 (1995).
2. Butler, P. & Marcy, G. Detection of extrasolar planets. *Annu. Rev. Astron. Astrophys.* (in the press).
3. Aumann, H. H. *et al.* Discovery of a shell around α Lyrae. *Astrophys. J.* **278**, L23–L27 (1984).
4. Smith, B. A. & Terrell, R. J. The circumstellar disk around β Pictoris. *Science* **226**, 1421–1424 (1984).
5. Kalas, P. & Jewitt, D. The detectability of β Pic-like circumstellar disks around nearby main-sequence stars. *Astron. J.* **111**, 1347–1355 (1996).
6. Zuckerman, B. & Becklin, E. E. Submillimetre studies of main-sequence stars. *Astrophys. J.* **414**, 793–802 (1993).

7. Chini, R., Krügel, E., Kreysa, E., Shustov, B. & Tutukov, A. Dust disks around Vega-type stars. *Astron. Astrophys.* **252**, 220–228 (1991).
8. Holland, W. S. *et al.* in *Advanced Technology MMW, Radio and Terahertz Telescopes* (ed Phillips, T.) (*Proc. SPIE* 3357, in the press).
9. Jenness, T. & Lightfoot, J. F. in *Starlink User Manual* Vol. 216 (1997).
10. Backman, D. E. & Paresce, F. in *Protostars and Planets III* (eds Levy, E. H. & Lunine, J. I.) 1253–1304 (Univ. Arizona Press, Tucson, 1993).
11. Lester, D., Harvey, P., Smith, B., Colomé, C. & Low, F. The far-IR size of the dust cloud around Fomalhaut. *Bull. Am. Astron. Soc.* **21**, 1085–1086 (1989).
12. Backman, D. E., Dasgupta, A. & Stencil, R. E. Model of a Kuiper Belt small grain population and resulting far-IR emission. *Astrophys. J.* **450**, L35–L38 (1995).
13. Harvey, P. M., Smith, B. J., DiFrancesco, J., Colomé, C. & Low, F. J. Far-IR constraints on dust shells around Vega-like stars. *Astrophys. J.* **471**, 973–978 (1996).
14. Jura, M. *et al.* A proto-cometary cloud around HR 4796A? *Astrophys. J.* (submitted).
15. Barrado y Navascués, D., Stauffer, J. R., Hartmann, L. & Balachandran, S. C. The age of Gliese 879 and Fomalhaut. *Astrophys. J.* **475**, 313–321 (1997).
16. Nakano, T. Formation of planets around stars of various masses—II. Stars of two and three solar masses and the origin and evolution of interstellar dust clouds. *Mon. Not. R. Astron. Soc.* **230**, 551–571 (1988).
17. Armitage, P. J. & Pringle, J. E. Radiation-induced warping of protostellar accretion disks. *Astrophys. J.* **488**, L47–L50 (1997).
18. Kalas, P. & Jewitt, D. Asymmetries in the β Pictoris disk. *Astron. J.* **110**, 794–804 (1995).
19. van der Blik, N. S., Prusti, T. & Waters, L. B. F. M. Vega: smaller dust grains in a larger shell. *Astron. Astrophys.* **285**, 229–232 (1994).
20. Gulliver, A. F., Hill, G. & Adelman, S. J. Vega: a rapidly rotating pole-on star. *Astrophys. J.* **429**, L81–L84 (1994).
21. Heap, S. R. *et al.* Hubble/STIS coronagraphic imagery of β Pictoris. *Bull. Am. Astron. Soc.* **29**, 1285 (1997).
22. Maeder, A. & Meynet, G. Tables of evolutionary star models from 0.85 to 120 solar masses, with overshooting and mass loss. *Astron. Astrophys. Suppl.* **76**, 411–425 (1988).
23. Mannings, V. G. & Emerson, J. P. Dust in disks around T-Tauri stars: Grain growth? *Mon. Not. R. Astron. Soc.* **267**, 361–378 (1994).
24. Burrows, A. *et al.* in *Brown Dwarfs and Extrasolar Planets* (eds Rebolo, R., Martin, E. L. & Zapatero-Osorio, M. R.) 354–369 (ASP conf. ser. Vol. 134, 1998).
25. Maher, K. A. & Stevenson, D. J. Impact frustration of the origin of life. *Nature* **331**, 612–614 (1988).
26. Pollack, J. B. *et al.* Composition and radiative properties of grains in molecular clouds and accretion disks. *Astrophys. J.* **421**, 615–639 (1994).
27. Freudling, W., Lagrange, A.-M., Vidal-Madjar, A., Ferlet, R. & Forveille, T. Gas around β Pictoris: an upper limit on the HI content. *Astron. Astrophys.* **301**, 231–235 (1995).
28. Smail, I., Ivison, R. J. & Blain, A. W. A deep submillimetre survey of lensing clusters: A new window on galaxy formation and evolution. *Astrophys. J.* **490**, L5–L8 (1997).
29. Jura, M. *et al.* The fate of the solid matter orbiting HR4796A. *Astrophys. J.* **445**, 45–53 (1995).
30. Mannings, V. G. Submillimetre observations of Herbig Ae/Be systems. *Mon. Not. R. Astron. Soc.* **271**, 587–600 (1994).

Acknowledgements. The JCMT is operated by the Joint Astronomy Centre, on behalf of the UK Particle Physics and Astronomy Research Council, the Netherlands Organization for Scientific Research, and the Canadian National Research Council. This research was also supported, in part, by NSF and NASA grants to UCLA. We thank E. Becklin for suggestions, and M. Plavec for information regarding stellar models. We also thank P. Kalas and B. A. Smith for sharing unpublished data with us.

Correspondence and requests for materials should be addressed to W.S.H. (e-mail: wsh@jach.hawaii.edu).

Shallow mixing in the solar photosphere inferred from revised beryllium abundances

Suchita C. Balachandran & Roger A. Bell

Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA

The chemical compositions of the Sun and meteorites are the benchmarks against which the abundances of elements in all other astronomical objects are compared. A long-standing problem¹ has been the abundance of lithium in the Sun's photosphere, which is ~ 140 times less than the meteoritic value (which represents the lithium abundance at the time the Solar System formed). This depletion requires that material from the photosphere be transported below the convective zone into regions where the temperature is high enough that nuclear processing can remove lithium. The models^{2,3} best able to do so simultaneously deplete beryllium by about a factor of two, which is consistent with previous measurements^{4,5} of the beryllium abundance in the solar photosphere. But here we show that these previous measurements are in error, because they did not fully account for the continuous opacity in the ultraviolet region of the spectrum where the beryllium lines are observed. We find that, after

correcting for this opacity, solar beryllium is not depleted at all with respect to the meteoritic value. This implies that mixing in the solar photosphere is more superficial than had hitherto been supposed, consistent with the shallow mixing inferred from recent helioseismic data⁶.

The only lines available for ground-based Be measurements are the Be II resonance lines at 3,130.42 and 3,131.06 Å near the atmospheric ultraviolet (UV) cut-off. In addition to the analytical difficulty of obtaining accurate abundances from crowded spectral regions, a singular problem plagues UV measurements: the uncertain UV opacity. An increase in this opacity is demanded by theoretical models of the solar atmosphere which consistently overpredict the UV flux⁷. Although the suggestion has been made that this mismatch is due to the insufficient inclusion of line opacity⁸, comparisons of high-resolution solar and synthetic spectra have shown that when all of the spectral features are reasonably well accounted for, the calculated solar flux remains too large⁹. Thus, the 'missing UV opacity' must be due to a continuous opacity or to a very large number of very weak lines¹⁰ which essentially mimic a continuous opacity.

To quantify the 'missing UV opacity' in the solar models, we have taken the empirical approach of requiring the UV A–X electronic transitions of the OH molecule to yield the same oxygen abundance as the vibrational–rotational OH transitions in the infrared (IR). The IR transitions, regarded as one of the more reliable oxygen-abundance indicators¹¹, are used to obtain an accurate oxygen abundance in the Sun. The oscillator strengths of the UV lines are accurately known, with experimental and theoretical values agreeing to within 10% (ref. 12). If the continuous opacity in the Sun is not fully accounted for in the synthetic spectral calculations, the calculated UV OH features will be stronger than observed. The mismatch between the observed and predicted line strengths is attributable to the 'missing UV opacity'.

We used the high-resolution ($\Delta\lambda = 348,000$) solar flux spectral atlas¹³ for our investigation. We searched an 80-Å region of the solar spectrum from 3,100 to 3,180 Å for clean (unblended) OH features. Fourteen features satisfied our condition that the sole contribution to the synthetic line comes from the OH feature. The observed feature may, of course, contain blends not identified in our list of lines⁹. Our empirical process therefore produces a lower limit to the 'missing' opacity. Three model solar atmospheres were used in our analysis: the Holweger–Müller (HM) empirical model¹⁴, the Kurucz solar model without overshoot¹⁵ and the updated OSMARCS model¹⁶. As OH is a trace species in the solar atmosphere, the oxygen abundance derived from OH lines is sensitively dependent on the model characteristics. Using the solar equivalent widths for the IR vibrational–rotational OH lines¹⁷, the following values were obtained for the oxygen abundance on a logarithmic scale where the abundance of hydrogen is 12: $\log(\text{O}) = 8.91 \pm 0.02$ (HM), 8.80 ± 0.02 (Kurucz) and 8.75 ± 0.02 (OSMARCS). Synthetic spectra of UV OH lines generated from each model atmosphere and its corresponding IR-based oxygen abundance produced lines of identical strength, indicating that the UV and IR OH lines are formed in the same part of the atmosphere and under similar conditions. Our differential use of the OH lines as a benchmark for the oxygen abundance therefore precludes the need to know the absolute value of the oxygen abundance with high accuracy.

Synthetic UV spectra were calculated under the condition of local thermodynamic equilibrium. The following opacities were included in the calculation of the synthetic solar spectrum: H_{bf} , H_{bf} , H_{ff} , H_{2bf} , H_{2ff} , H and He Rayleigh scattering, He_{ff} and electron scattering; here subscript bf indicates bound–free, and ff indicates free–free. Metal bound–free opacities were not used in our calculation. Opacity Project (an international collaboration to calculate atomic data for stellar opacity computations) calculations of Si I and Mg I bound–free opacities¹⁸ show that they are relatively small at optical depth $(\tau)(3,100) = 0.1$, being roughly 1% and 8.5% respectively of the

total opacity from the H and He contributions. The Fe I bound-free opacity is not well known. Calculations¹⁹ place it at 4% of the total whereas the contribution from CH (ref. 20) is 3%.

With the standard continuous opacity and the IR-based oxygen abundance, all 14 OH features were calculated to be stronger than observed. Twelve features matched the observed OH line strength when the continuous opacity was increased by factors of between 1.5 and 1.7. Two lines, which presumably contain blends not recognized in our line list, required much smaller opacity increases. The OH lines at 3,128.06 and 3,128.10 Å are the cleanest, with the nearest neighbouring lines 0.2 Å away. Using this feature, we estimate the required opacity increase to be a factor of 1.6 ± 0.15 over the standard value. All three solar model atmospheres required an identical increase in opacity to fit the OH features. An example of the fit to the 3,128 Å OH feature with the OSMARCS model is shown in Fig. 1.

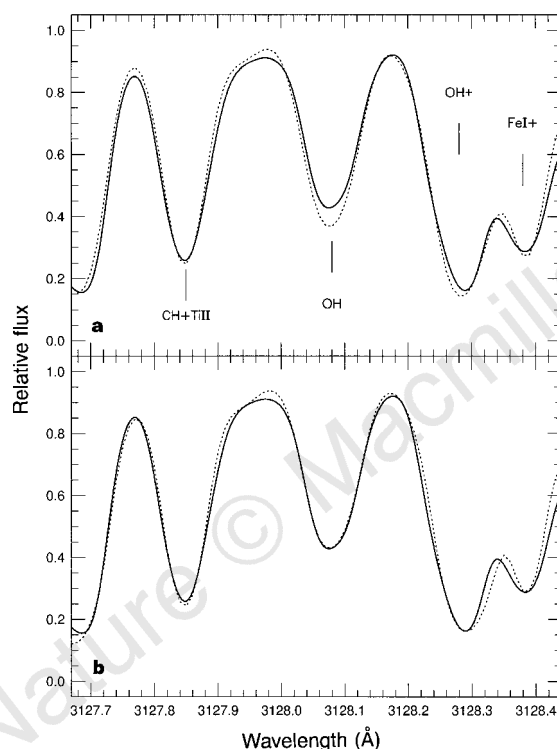


Figure 1 The solar flux spectrum¹³ around one clean OH feature (solid line) compared to calculated spectra (dotted lines). In **a**, a synthetic fit to the spectrum using the standard opacity is shown by the dotted line. To produce the synthesis, the OH wavelengths were taken from Stark *et al.*²⁵, and the Einstein *A* coefficients²⁶ were based on the lifetime measurement of 688 ns for the 0–0 ground state²⁷ which is in good agreement with the most recent theoretical calculation of 673 ns (ref. 12). The remaining features used in the initial construction of the synthetic spectrum were taken from the NVISLINES list⁸. As the oscillator strengths of the atomic features are not known to the accuracy of the OH lines, they were adjusted to match the observed spectrum. This synthesis was produced with the OSMARCS solar model and hence the oxygen abundance of $\log \epsilon(\text{O}) = 8.75$ was used. The clean OH feature at 3128.06 is marked. Other significant features are identified. OH+ and FeI+ are blended features which have OH and Fe I as principal components. In **b**, the standard continuous opacity has been multiplied by a factor of 1.6 to fit the OH feature. The observed feature may, of course, contain blends not identified in our line list. Our empirical process therefore produces a lower limit to the ‘missing’ opacity.

Alternative explanations of the lack of agreement with the standard opacity may be investigated. For example, the oscillator strengths of the OH features may have larger errors than claimed in the original studies. In such a case, it should be possible to match all the OH features by altering the oscillator strengths by a single factor. No such solution could be found. Changes in oscillator strength of -0.15 to -0.3 dex are required to match OH lines, all of which arise from the 0–0 band. Differential changes of such magnitude are disallowed for a given band strength. Furthermore, as such large differential changes are required for lines which have nearly the same lower excitation energy levels and which arise from the same region of the solar atmosphere, they cannot be attributed to inaccuracies in the temperature structure of the model atmosphere.

A synthetic spectrum of the Be II region produced under conditions of local thermal equilibrium and with the standard opacity

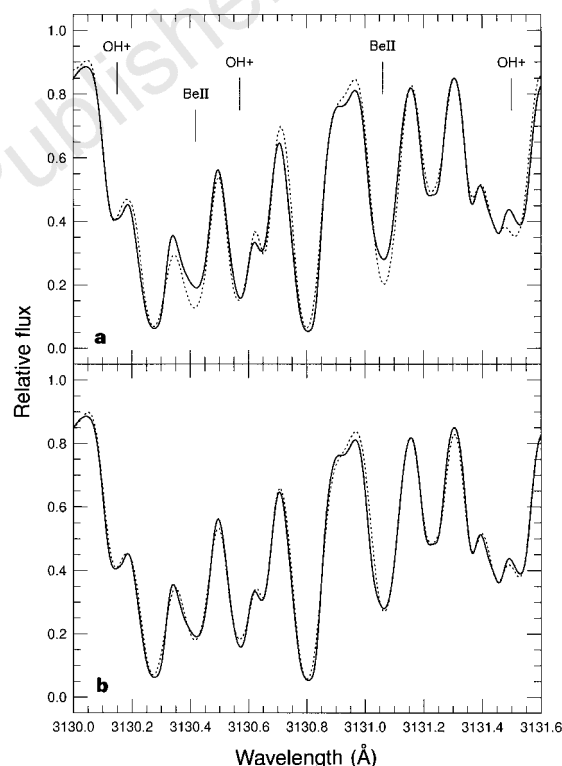


Figure 2 The Be II region in the solar flux spectrum¹³ (solid line) compared to calculated spectra (dotted lines). The Be II lines and blended OH lines (OH+) are indicated. Be II is not a trace species in the Sun. The Be II lines are therefore not sensitively dependent on the input model, in contrast to the OH lines. Using the empirical HM model as standard with oscillator strengths from ref. 28, we obtain $\log \epsilon(\text{Be}) = 1.42$ (HM), 1.40 (Kurucz) and 1.38 (OSMARCS) for the same input equivalent width. We have adopted these values as the undepleted photospheric Be abundance appropriate for each model atmosphere. The synthetic spectra were produced assuming local thermodynamic equilibrium (LTE), with $\log \epsilon(\text{Be}/\text{H}) = 1.36$ and the OSMARCS solar model¹⁶, and are shown as dotted lines in each panel. Non-LTE effects are calculated to be small^{29,30} for Be II. The uncertain oscillator strengths of the atmospheric features have been altered separately in each panel to fit the observed spectrum. Be II oscillator strengths²⁸ are not changed. The standard continuous opacity (in **a**) produces Be II lines which are stronger than the observed features. This fact has led previous researchers^{4,5} to conclude that the Sun has depleted its surface Be. However, when the continuous opacity is increased by a factor of 1.6, as required to reproduce the OH features, the Be II features are matched (in **b**). This leads to our conclusion that the Sun has retained its initial Be abundance.

and the meteoritic Be abundance is shown in Fig. 2a. In striking resemblance to the OH lines, the Be II lines are too strong. This illustrates why previous studies^{4,5} have concluded that Be must be depleted in the solar photosphere. But the synthetic spectrum with the continuous opacity increased by a factor of 1.6 is in very good agreement with the observed Be II lines (Fig. 2b). This leads to our conclusion that the Sun had not depleted any of its surface Be since its formation 4.6 Gyr ago. Combining the error in the opacity estimate and the Be II line fit, we obtain $\log \epsilon(\text{Be}) = 1.40 \pm 0.09$ relative to the initial photospheric (that is, meteoritic) abundance of $\log \epsilon(\text{Be}) = 1.42$.

Although the source of the 'missing opacity' must remain speculative until atomic calculations are improved, some insight may be gained from existing results. Early observations²¹ of the Be II lines at two points on the solar disk, near the limb ($\cos \theta = 0.2$, where θ is the angle made by a line from the solar centre to any point on the disk with the outward normal) and at the centre ($\cos \theta = 1.0$), showed that the line depths could not be matched unless the continuous opacity was increased. The observed ratio of the depth of the 3,131.06-Å Be II line at these two points on the solar disk is 0.70 (ref. 21). The synthetic spectrum with opacity increased by a factor of 1.6 yields a ratio of 0.89, which is only slightly smaller than the ratio of 0.90 derived from the standard-opacity case. We find that an increase in the Fe I bound-free opacity of a factor of 32 from the estimate we used¹⁹ mimics the factor of 1.6 increase in the continuous opacity in matching line strengths in the solar flux spectrum. Furthermore, as the Fe I bound-free opacity contributes predominantly to the outer layers of the atmosphere, the effect on the solar limb darkening is different and the predicted ratio of the Be II line depths at $\cos \theta = 0.2$ and 1.0 is 0.74, in close agreement with the observed ratio.

Preliminary calculations also show that an increase in the Fe I opacity greatly improves the agreement between the calculated (calc.) and observed (obs.) solar fluxes in the 3,130-Å region. For example, an increase of a factor of 32 gives (obs. - calc.)/obs. = -0.18 compared to the standard opacity case of (obs. - calc.)/obs. = -0.54 in the heavily blanketed 3,100–3,150 Å region; here the observational data is taken from the SOLSTICE²² experiment.

Given the uncertainty in the Fe I bound-free opacity, we particularly urge detailed calculations of this quantity, though other opacities may contribute as well. If the 'missing opacity' arises from a metal bound-free source, it will decrease as a function of metallicity and have an insignificant contribution in metal-poor stars. Although it is possible to check this effect by similar empirical estimates in stellar spectra, acquisition of data with the required high resolution and high signal-to noise ratio is not easy. Until such time, abundances derived from UV features must be regarded as uncertain, at least in solar-metallicity stars.

The observed depletion of Li clearly indicates the presence of mixing in the solar interior which is not accounted for by the standard solar model, because the temperature at the base of the Sun's surface convection zone is only 2×10^6 K whereas a temperature of 2.5×10^6 K is needed to destroy Li. The lack of Be depletion, which requires a temperature of 3.5×10^6 K, now provides a strong constraint on the depth to which turbulence is effective below the surface convection zone. Current turbulence models produce too much mixing. Using a slow mixing process³ which has been linked parametrically with the transport of angular momentum², such models predict the depletion of both Li and Be in the Sun. The agreement that we find between the solar and meteoritic Be abundance argues otherwise, and implies that mixing is more superficial than previously thought.

Our finding is also consistent with recent helioseismic data which indicates the existence of only a thin shear layer below the surface convection zone in the present Sun⁶, as may be required for Li, but perhaps not Be, destruction. With the advance in helioseismological techniques which are able to probe the solar interior with unre-

cedented sensitivity, it will soon be possible to fully constrain the important region below the solar convection zone.

Whether or not the transport of material is linked to the transport of angular momentum remains an important unanswered question. Angular-momentum transport is a fundamental process needed to explain how the angular velocity gained during gas collapse and star formation is lost once the star is on the main sequence. The strong constraint on mixing depth imposed by the lack of solar Be depletion must now be met by revised transport models. If the turbulence-related angular-momentum transport models are unable to do so, the link between angular momentum and particle transport may be more tenuous than assumed. Indeed, alternative suggestions for the transport of angular momentum by internal gravity waves—which leave the young Sun rotating as a solid body within 10 Myr—have recently been proposed^{23,24}.

The entire problem of the abundance of Li in the Sun, which has puzzled astronomers for over four decades, requires further close scrutiny. It is only after the mechanism of Li depletion is fully understood that the primordial Li abundance can be calculated from measurements of the oldest metal-poor stars in the Galaxy. □

Received 20 November 1997; accepted 11 February 1998.

- Pinsonneault, M. H. Mixing in stars. *Annu. Rev. Astron. Astrophys.* **35**, 557–605 (1997).
- Pinsonneault, M. H., Kawaler, S. D., Sofia, S. & Demarque, P. Evolutionary models of the rotating sun. *Astrophys. J.* **338**, 424–452 (1989).
- Richard, O., Vauclair, S., Charbonnel, C. & Dziembowski, W. A. New solar models including helioseismological constraints and light-element depletion. *Astron. Astrophys.* **312**, 1000–1011 (1996).
- King, J. R., Deliyannis, C. P. & Boesgaard, A. M. The ⁷Be abundances of alpha centauri A and B and the sun: Implications for stellar evolution and mixing. *Astrophys. J.* **478**, 778–786 (1997).
- Primas, F., Duncan, D. K., Pinsonneault, M. H., Deliyannis, C. P. & Thorburn, J. A. Hubble Space Telescope beryllium abundances in the alpha centauri system. *Astrophys. J.* **480**, 784–793 (1997).
- Basu, S. Seismology at the base of the solar convective zone. *Mon. Not. R. Astron. Soc.* **288**, 572–584 (1997).
- Gustafsson, B. & Bell, R. A. The colours of G and K type giant stars. I. *Astron. Astrophys.* **74**, 313–352 (1979).
- Kurucz, R. L. Finding the "missing" solar ultraviolet opacity. *Rev. Mex. Astron. Astrofis.* **23**, 181–194 (1992).
- Bell, R. A., Paltoglou, G. & Tripicco, M. J. The calibration of synthetic colours. *Mon. Not. R. Astron. Soc.* **268**, 771–792 (1994).
- Holweger, H. Depression of the solar continuous spectrum by line absorption and Balmer continuum. *Astron. Astrophys.* **4**, 11–17 (1970).
- Grevesse, N. & Sauval, A. J. in *Molecules in the Stellar Environment* (ed. Jorgensen, U. G.) 196–208 (IAU Colloq. 146, Springer, Berlin, 1994).
- Bauschlicher, C. W. Jr & Langhoff, S. R. Theoretical determination of the radiative lifetime of the A²Σ⁺ state of OH. *J. Chem. Phys.* **87**, 4665–4672 (1987).
- Kurucz, R. L., Furenlid, I., Brault, J. & Testerman, L. *Solar Flux Atlas from 296 to 1200 nanometers* (Nat'l Solar Observatory, Tucson, 1984).
- Holweger, H. & Müller, E. A. The photospheric barium spectrum—Solar abundance and collisional broadening of Ba II lines by hydrogen. *Solar Phys.* **39**, 19–30 (1974).
- Kurucz, R. L. *Atlas 9 Stellar Atmosphere Programs and 2cm/s Grid CD-ROM No. 13* (Smithsonian Astrophysical Observatory, 1993).
- Edvardsson, B. *et al.* The chemical evolution of the galactic disk—part one—analysis and results. *Astron. Astrophys.* **275**, 101–152 (1993).
- Grevesse, N., Sauval, A. J. & van Dischoeck, E. F. An analysis of vibration-rotation lines of OH in the solar infrared spectrum. *Astron. Astrophys.* **141**, 10–16 (1984).
- Seaton, M. J., Yan, Y., Mihalas, D. & Pradhan, A. K. Opacities for stellar envelopes. *Mon. Not. R. Astron. Soc.* **266**, 805–828 (1994).
- Dragon, J. N. & Mutschlechner, J. P. The Solar Ultraviolet Continuum. *Astrophys. J.* **239**, 1045–1069 (1980).
- Kurucz, R. L., van Dischoeck, E. F. & Tarafdar, S. P. OH and CH continuous opacity in solar and stellar atmospheres. *Astrophys. J.* **322**, 992–998 (1987).
- Chmielewski, Y., Müller, E. A. & Brault, J. W. The solar beryllium abundance. *Astron. Astrophys.* **42**, 37–46 (1975).
- Woods, T. N. *et al.* Validation of the UARS solar ultraviolet irradiances: Comparison with the ATLAS1 and 2 measurements. *J. Geophys. Res.* **D6**, 9541–9569 (1996).
- Kumar, P. & Quataert, E. J. Angular momentum transport by gravity waves and its effect on the rotation of the solar interior. *Astrophys. J.* **475**, L143–L146 (1997).
- Zahn, J.-P., Talon, S. & Matias, J. Angular momentum transport by internal waves in the solar interior. *Astron. Astrophys.* **322**, 320–328 (1997).
- Stark, G., Brault, J. W. & Abrams, M. C. Fourier-transform spectra of the A²Σ⁺ - X²II Δ*v* = 0 bands of OH and OD. *J. Opt. Soc. Am. B* **11**, 3–32 (1994).
- Goldman, A. & Gillis, J. R. Spectral line parameters for the A²Σ⁺ - X²II(0,0) band of OH for atmospheric and high temperatures. *J. Quant. Spectrosc. Radiat. Transfer* **25**, 111–135 (1981).
- German, K. R. Direct measurement of the radiative lifetimes of the A²Σ⁺ (*V* = 0) states of OH and OD. *J. Chem. Phys.* **62**, 2584–2587 (1974).
- Wiese, W. L. & Martin, G. A. *Wavelengths and Transition Probabilities for Atoms and Atomic Ions* (NBS Monogr. 68, US Govt Print. Off., Washington, DC, 1980).
- García López, R. J., Severino, G. & Gomez, M. T. Galactic evolution of beryllium. I. NLTE effects and accuracy of beryllium abundances in metal-poor stars. *Astron. Astrophys.* **297**, 787–801 (1995).
- Kiselman, D. & Carlsson, M. in *The Light Element Abundances* (ed. Crane, P.) 372–381 (Proc. ESO/EPIC Workshop, Springer, Berlin, 1995).

Acknowledgements. We thank C. Proffitt and D. Lambert for discussion. This work was supported by a NASA grant to S.C.B. and a NSF grant to R.A.B.

Correspondence and requests for materials should be addressed to S.C.B. (e-mail: suchitra@astro.umd.edu).

Direct evidence for a half-metallic ferromagnet

J.-H. Park*, E. Vescovo*, H.-J. Kim*, C. Kwon†‡, R. Ramesh† & T. Venkatesan†

* NSLs Brookhaven National Laboratory, Upton, New York 11973-5000, USA

† Center for Superconductivity Research, Department of Physics, University of Maryland, College Park, Maryland 20742, USA

Half-metallic materials are characterized by the coexistence of metallic behaviour for one electron spin and insulating behaviour for the other. Thus, the electronic density of states is completely spin polarized at the Fermi level, and the conductivity is dominated by these metallic single-spin charge carriers. This exotic physical property could have a significant effect on technological applications related to magnetism and spin electronics. Some ferromagnetic systems, such as Heusler compounds¹ and chromium dioxide², have been predicted theoretically to be half-metallic. However, a half-metallic system has not been demonstrated directly and the predictions are still in doubt^{3,4}. Here we report spin-resolved photoemission measurements of a ferromagnetic manganese perovskite, $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$, which directly manifest the half-metallic nature well below the Curie temperature. For the majority spin, the photoemission spectrum clearly shows a metallic Fermi cut-off, whereas for the minority spin, it shows an insulating gap with disappearance of spectral weight at ~ 0.6 eV binding energy.

Half-metallic ferromagnets were introduced by de Groot *et al.*¹, who predicted a metallic character for majority-spin electrons but an insulating character—that is, the Fermi energy (E_F) falls in a gap—for minority-spin electrons, in the spin-polarized band-structure calculation for a Mn-based Heusler alloy NiMnSb ². The half-metallic ferromagnet has 100% spin polarization for the conduction electrons, and thus it offers potential technological applications such as a single-spin electron source, and high-efficiency magnetic sensors. The half-metallic feature has also been predicted² for a prototypical ferromagnetic oxide CrO_2 . However, spin-resolved photoemission studies have not confirmed these predictions. In the case of NiMnSb , only $\sim 50\%$ spin polarization was reported for the conduction electrons³. For CrO_2 , $\sim 100\%$ spin polarization was observed at ~ 2 eV binding energy rather than at E_F , where almost no spectral weight appears for both spin electrons, and even the metallic character was questioned⁴. The photoemission measurements are rather surface sensitive, and it is known that it is quite difficult to obtain proper stoichiometry at the surface of both these systems. Thus, it has been often suggested that the results might be influenced by surface segregation effects^{3,5}.

Doped manganese perovskites with a formula $\text{Ln}_{1-x}\text{A}_x\text{MnO}_3$ (where Ln is a trivalent rare-earth element and A is a divalent alkaline-earth element) have attracted considerable interest, particularly with reference to their potential technological applications. On cooling, these compounds show a large decrease in resistivity associated with a paramagnetic to ferromagnetic transition^{6,7}. This transition is known to give rise to a large negative magnetoresistance, the so-called colossal magnetoresistance (CMR)^{8,9}, near the Curie temperature, T_C . The ferromagnetic ground state has been explained by the double-exchange model¹⁰. However, a model considering primarily the spin-dependent electron hopping mechanism turns out not to be able to explain the insulator-like high resistivity above T_C . Recent theoretical and experimental studies indicate that small polaron effects (including the Jahn–

Teller distortion), as well as the double-exchange model, are required for understanding of the transition and the transport measurements^{11–13}.

The doped manganese perovskite is mixed-valent with Mn^{3+} ($3d^4$) and Mn^{4+} ($3d^3$). For the octahedral site symmetry of the MnO_6 complex, the configuration becomes $t_{2g}^3e_g^1$ (5E_g) for Mn^{3+} and t_{2g}^3 (4A_g) for Mn^{4+} . In the double-exchange mechanism, the e_g electrons are considered as mobile charge carriers interacting with the localized Mn^{4+} ($S = 3/2$) spins. The carrier hopping avoids the strong on-site Hund rule exchange energy J_{ex} when the Mn spins are aligned ferromagnetically. J_{ex} (~ 2.5 eV) is much larger than the e_g band width (~ 1 eV). Thus the conduction electrons are expected to be highly spin polarized, and the possibility of 100% spin polarization (that is, a half-metallic state) has been discussed^{14–16}. However, a band-structure calculation¹⁷ of $\text{La}_{2/3}\text{Ca}_{1/3}\text{MnO}_3$ predicted $\sim 36\%$ spin polarization at E_F , and previous spin-resolved tunnelling measurements reported 54% (ref. 18) and 81% (ref. 19) polarization for the conduction electrons of $\text{La}_{0.67}\text{Sr}_{0.33}\text{MnO}_3$ tunnel junctions.

We grew a 1,900-Å-thick $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ epitaxial thin film on a (001) SrTiO_3 single-crystal substrate by conventional pulse laser deposition. The surface roughness of the sample is expected to be ~ 20 Å, and the T_C was determined to be ~ 350 K. The details of sample growth and characterization are described elsewhere²⁰. The measurements were performed at the new U5UA undulator beamline at the National Synchrotron Light Source (NSLS) in the Brookhaven National Laboratory²¹. During the measurements, the base pressure was kept better than 1×10^{-10} torr. The sample was introduced into a vacuum, and the surface was cleaned *in situ* in a sequence of annealing processes, which will be presented in detail elsewhere. The T_C at the surface turns out to be nearly identical to that of the film, indicating that the deviation of the oxygen stoichiometry at the surface is negligible. The photoelectrons were collected by an electron analyser with $\pm 2^\circ$ angular resolution. However, the spectrum does not display any noticeable angular dependence, indicating that the momentum ($\mathbf{k}$) resolution in reciprocal space, which requires a well defined surface normal, is completely obscured owing to the rough surface. Thus, the results obtained can be considered as angle-integrated.

Successful achievement of a clean surface was confirmed in our measurements of temperature-dependent high-resolution photoemission. Figure 1 shows the spectra near E_F measured at two different temperatures, $T = 40$ K ($\ll T_C$) and $T = 380$ K ($> T_C$). The electronic change through the ferromagnetic transition is clearly visible: at $T = 380$ K, the spectrum shows negligible spectral weight at E_F , whereas at $T = 40$ K it clearly shows the metallic Fermi cut-off. In detailed temperature-dependence studies, we also

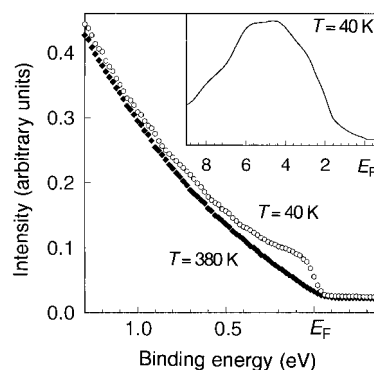


Figure 1 High-resolution photoemission spectra of a thin film of $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$, very near the Fermi energy (E_F). The T_C of the sample was ~ 350 K, measurements were taken at $T = 40$ K ($\ll T_C$; **a**) and at $T = 380$ K ($> T_C$; **b**). The photon energy and the experimental resolution were set at $h\nu = 40$ eV, and $\Delta E = 0.1$ eV, respectively. The inset shows the wide-scan valence band spectrum at $T = 40$ K.

‡ Present address: Superconductivity Technology Center, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.

observed an increase of the spectral weight at E_F with cooling below T_C . This temperature dependency is now well known to be one of the peculiar characteristics of the doped manganese perovskites^{12,22}. The wide-range spectrum shown in Fig. 1 inset is also very similar to that reported for bulk polycrystalline $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ samples²². The electronic state above T_C of $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ ($0.17 < x < 0.5$) is not yet well characterized; the temperature dependence of the resistivity of such compounds becomes more metallic-like with increasing of the Sr concentration x , in spite of their high resistance above T_C (ref. 7). However, the spectroscopic measurements show only negligible spectral weight at E_F without the metallic feature, and the high-temperature state has been suggested to be a pseudogap state^{12,22,23}.

The electrical and magnetic properties of $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ thin

films are known to be nearly identical to those of the bulk samples^{7,20,24}, but their magnetic domain behaviour turns out to be completely different from that of the bulk, possibly because of surface stress and finite-size effects. The bulk crystals typically required fields of several teslas to saturate the magnetization and yielded very little remanent magnetization. However, as shown in Fig. 2a, inset, the $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ thin-film sample shows a single-domain-type magnetic hysteresis, that is $\sim 100\%$ remanent magnetization and low coercive field (~ 100 Oe at 40 K). This single-domain-type behaviour is crucial in opening up the possibility of spin-resolved photoemission studies.

Figure 2 shows valence band spin-resolved photoemission spectra near E_F of the $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ thin film that we measured at

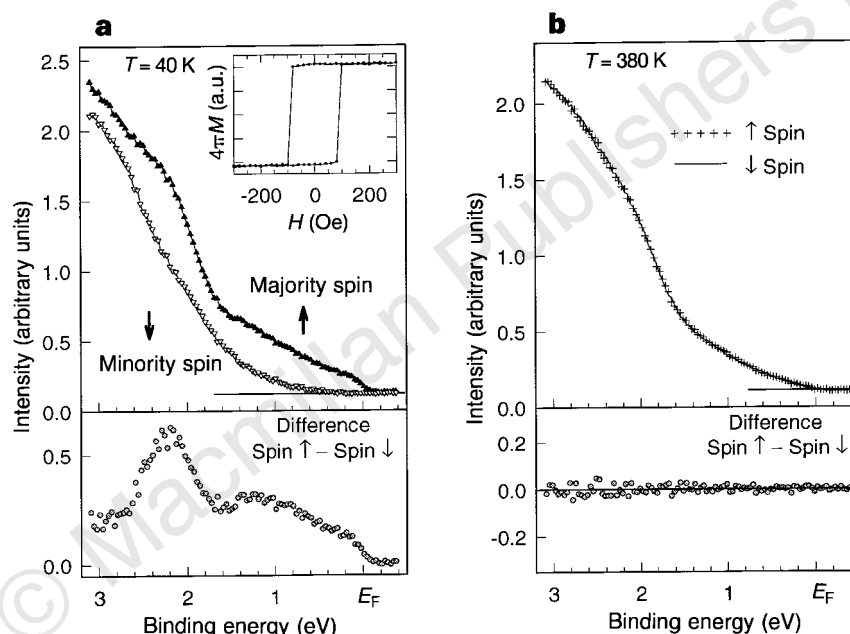
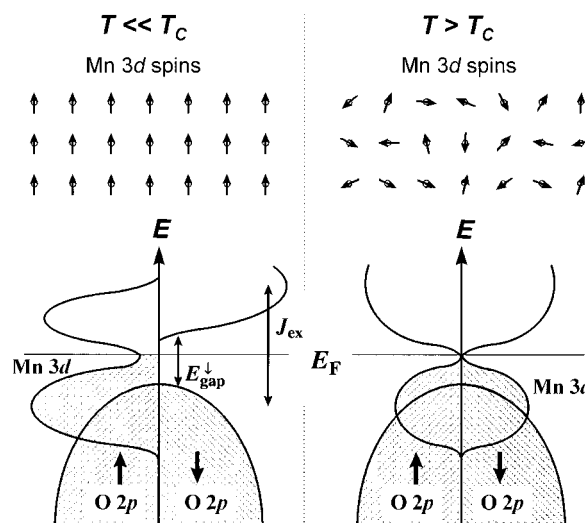


Figure 2 Spin-resolved photoemission spectra of a thin film of $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ (as in Fig. 1), near the Fermi energy (E_F). Measurements were taken at $T = 40$ K ($\ll T_C$; **a**) and at $T = 380$ K ($> T_C$; **b**). The photon energy and the experimental resolution were set at $h\nu = 40$ eV, and $\Delta E = 0.2$ eV, respectively. A magnetic pulse coil with ~ 200 Oe magnetic field was used for magnetization of the sample. The majority ($\uparrow$) and minority ($\downarrow$) spins represent the spin directions respectively parallel

and anti-parallel to the magnetization direction. The bottom panels of **a** and **b** show the difference spectra between the majority-spin and the minority-spin experimental spectra. The inset in **a** shows the magnetization (M) versus applied magnetic field (H) hysteresis loop, which was obtained by monitoring Mn L_2 -edge absorption intensity for circularly polarized incident light at the NSLS U4B beamline.

Figure 3 Schematic energy diagrams and the Mn 3d spin alignments of the doped manganese perovskites at $T \ll T_C$ and at $T > T_C$. J_{ex} is the Hund rule exchange energy, and $E_{\text{gap}\downarrow}$ denotes the insulating bandgap of the minority-spin states.



$T = 40$ K (Fig. 2a) and $T = 380$ K (Fig. 2b). At $T = 40$ K ($\ll T_C$), all the Mn 3d electron spins are aligned ferromagnetically. Indeed, the spectra in Fig. 2a show considerable difference for the majority ($\uparrow$) and minority ($\downarrow$) spins. The most striking observation is the half-metallic feature. The spectrum for the majority spin extends up to E_F and shows the metallic Fermi cut-off, while that for the minority spin decreases rapidly at ~ 1 eV binding energy and the spectral weight disappears very near E_F reflecting the insulating gap. In the region of 0 to ~ 0.4 eV binding energy, the minority-spin states only show spectral noise which is also present above E_F and the spin polarization is found to be $100 \pm 5\%$. Considering the spectral broadening due to the finite experimental resolution, we estimate the half-gap on the occupied state (photoemission) side of the minority-spin states to be 0.6 ± 0.2 eV.

The minority-spin states still show large spectral weight in the higher-binding-energy region owing to the O 2p states, which are fully occupied for both spins. The Mn 3d state spectrum can be obtained by subtracting the minority-spin spectrum from the majority-spin spectrum. The difference spectrum presented in the bottom panel of Fig. 2a shows the metallic Fermi cut-off at E_F and two peak features around 1.2 eV and 2.2 eV binding energies, which can be interpreted as the e_g and t_{2g} electron removal states, respectively¹². A similar line shape was previously observed in Mn L_{3-2} -edge on-resonant photoemission spectra of bulk $\text{La}_{0.67}\text{Ca}_{0.33}\text{MnO}_3$ and $\text{La}_{0.7}\text{Pb}_{0.3}\text{MnO}_3$, in which the Mn 3d states are enhanced exclusively¹².

On heating through T_C , the Mn 3d spins become disordered, and the spin anisotropy disappears. The spin-resolved photoemission spectra above T_C confirm the disappearance of the spin anisotropy. The spectra in Fig. 2b, which were measured at $T = 380$ K ($> T_C$), show no difference for the two different spins, and the spectral weight disappears at E_F for both spins, showing the pseudogap feature. The difference spectrum of the up ($\uparrow$) spin and down ($\downarrow$) spin spectra, presented in the bottom panel of Fig. 2b, exhibits only experimental noise.

Our spin-resolved photoemission studies of $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ show (1) the half-metal feature well below T_C and (2) the occurrence of the metal to pseudogap state transition accompanied by loss of the ferromagnetic order on heating through T_C . These results are summarized schematically in Fig. 3. At $T \ll T_C$, the Mn 3d electron spins are aligned ferromagnetically owing to the ferromagnetic double-exchange interaction. Thus, the Mn 3d states, which extend up to E_F have only the majority spin. However, the O 2p states, which have both majority and minority spins, only reach to ~ 0.6 eV below E_F . Therefore, the majority-spin states show the metallic feature with charge carriers from Mn 3d states, whereas the minority-spin states exhibit an insulating gap between the occupied O 2p states and the unoccupied Mn 3d minority-spin states. At $T > T_C$, the Mn 3d spins become disordered and the spin anisotropy disappears. According to the double-exchange mechanism, the loss of ferromagnetic order reduces the electron hopping energy and the density of states at E_F almost disappears; this results from the competition of the electron hopping energy and the small polaron effects, including the Jahn–Teller distortion^{11–13}. □

Received 18 June 1997; accepted 27 January 1998.

- de Groot, R. A., Muller, F. M., van Engen, P. G. & Buschow, K. H. J. New class of materials: half-metallic ferromagnets. *Phys. Rev. Lett.* **50**, 2024–2027 (1983).
- Schwarz, K. CrO₂ predicted as a half-metallic ferromagnet. *J. Phys. F* **16**, L211–L215 (1986).
- Bona, G. L., Meier, F., Taborelli, M., Bucher, E. & Schmidt, P. H. Spin polarized photoemission from NiMnSb. *Solid State Commun.* **56**, 391–394 (1985).
- Kämper, K. P., Schmitt, W., Güntherodt, G., Gambin, R. J. & Ruf, R. CrO₂—a new half-metallic ferromagnet? *Phys. Rev. Lett.* **59**, 2788–2791 (1988).
- Hanssen, K. E. H. M., Mijnen, P. E., Rabou, L. P. L. M. & Buschow, K. H. J. Positron-annihilation study of the half-metallic ferromagnet NiMnSb: experiment. *Phys. Rev. B* **42**, 1533–1540 (1990).
- Jonker, G. H. & van Santen, J. H. Ferromagnetic compounds of manganese with perovskite structure. *Physica* **16**, 337–349 (1950).
- Tokura, Y. et al. Giant magnetotransport phenomena in filling-controlled Kondo lattice system: $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *J. Phys. Soc. Jpn* **63**, 3931–3935 (1994).
- von Helmolt, R., Wecker, J., Holszapfel, B., Schultz, L. & Samwer, K. Giant negative magnetoresistance in perovskitelike $\text{La}_{2/3}\text{Ba}_{1/3}\text{MnO}_x$ ferromagnetic films. *Phys. Rev. Lett.* **71**, 2331–2334 (1993).

- Jin, S. et al. Thousandfold change in resistivity in magnetoresistive La–Ca–Mn–O films. *Science* **264**, 413–415 (1994).
- Zener, C. Interaction between the d-shells in the transition metals. II. Ferromagnetic compounds of manganese with perovskite structure. *Phys. Rev.* **82**, 403–405 (1951).
- Millis, A. J., Shraiman, B. I. & Mueller, R. Dynamic Jahn–Teller effect and colossal magnetoresistance in $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. Lett.* **77**, 175–178 (1996).
- Park, J.-H. et al. Electronic aspects of the ferromagnetic transition in manganese perovskites. *Phys. Rev. Lett.* **76**, 4215–4218 (1996).
- de Teresa, J. M. et al. Evidence for magnetic polarons in the magnetoresistive perovskites. *Nature* **386**, 256–258 (1997).
- Okimoto, Y. et al. Anomalous variation of optical spectra with spin polarization in double exchange ferromagnet: $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. Lett.* **75**, 109–112 (1995).
- Hwang, H. Y., Cheong, S.-W., Ong, N. P. & Batlogg, B. Spin-polarized intergrain tunneling in $\text{La}_{2/3}\text{Sr}_{1/3}\text{MnO}_3$. *Phys. Rev. Lett.* **77**, 2041–2044 (1996).
- de Boer, P. K., van Leuken, H., de Groot, R. A., Rojo, T. & Barberis, G. E. Electronic structure of $\text{La}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$. *Solid State Commun.* **102**, 621–626 (1997).
- Pickett, W. E. & Singh, D. J. Electronic structure and half-metallic transport in the $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$ system. *Phys. Rev. B* **53**, 1146–1160 (1996).
- Lu, Y. et al. Large magnetotunneling effect at low magnetic fields in micrometer-scale epitaxial $\text{La}_{0.67}\text{Sr}_{0.33}\text{MnO}_3$ tunnel junctions. *Phys. Rev. B* **54**, R8357–R8360 (1996).
- Sun, J. Z., Krusin-Elbaum, L., Duncombe, P. R., Gupta, A. & Laibowitz, R. B. Spin-polarized tunneling in doped perovskite manganate trilayer junctions. *Appl. Phys. Lett.* **70**, 1769–1771 (1997).
- Kwon, C. et al. Stress-induced effects in epitaxial $(\text{La}_{0.7}\text{Sr}_{0.3})\text{MnO}_3$ films. *J. Magn. Magn. Mater.* **172**, 229–236 (1997).
- Vescovo, E. et al. The USUA SGM-beamline: First Tests on the Performances A-25 (Activity Rep. 1996, National Synchrotron Light Source, Brookhaven National Lab, Upton, 1997).
- Sarma, D. D. et al. Temperature-dependent photoemission spectral weight in $\text{La}_{0.6}\text{Sr}_{0.4}\text{MnO}_3$. *Phys. Rev. B* **53**, 6873–6877 (1996).
- Saitoh, T. et al. $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ studied by photoemission and X-ray absorption spectroscopy. *Phys. Rev. B* **51**, 13942–13951 (1995).
- Trajanovic, Z. et al. Growth of colossal magnetoresistance thin films on silicon. *Appl. Phys. Lett.* **69**, 1005–1007 (1996).

Acknowledgements. We thank J. Hastings for discussions. NSLS is supported by the US Department of Energy; the work in the University of Maryland was partially supported by the US Office of Naval Research.

Correspondence and requests for materials should be addressed to J.H.P. (e-mail: jhpark@bnlls1.nsls.bnl.gov).

A soft magnetic CoNiFe film with high saturation magnetic flux density and low coercivity

Tetsuya Osaka*, Madoka Takai*, Katsuyoshi Hayashi*, Keishi Ohashi†, Mikiko Saito‡ & Kazuhiko Yamada‡

* Department of Applied Chemistry, Waseda University, and Kagami Memorial Laboratory for Material Science and Technology, Waseda University, Shinjuku-ku, Tokyo 169, Japan

† NEC Ibaraki Ltd, Sekijomachi, Makabegun, Ibaraki 308-01, Japan

‡ NEC Corporation, Fuchu, Tokyo 183, Japan

Magnetic materials are classed as ‘soft’ if they have a low coercivity (the critical field strength H_c required to flip the direction of magnetization). Soft magnetic materials are a central component of electromagnetic devices such as step motors, magnetic sensors, transformers and magnetic recording heads. Miniaturization of these devices requires materials that can develop higher saturation flux density, B_s , so that the necessary flux densities can be preserved on reducing device dimensions, while simultaneously achieving a low coercivity. Common high- B_s soft magnetic films currently in use are electroplated CoFe-based alloys^{1–4}, electroplated CoNiFe alloys^{5–7}, and sputtered Fe-based nanocrystalline^{8–11} and FeN films^{12–14}. Sputtering is not suitable, however, for fabricating the thick films needed in some applications, for which electrochemical methods are preferred. Here we report the electrochemical preparation of a CoNiFe film with a very high value of B_s (2.0–2.1 T) and a low coercivity. The favourable properties are achieved by avoiding the need for organic additives in the deposition process, which are typically used to reduce internal stresses. Our films also undergo very small magnetostriction, which is essential to ensure that they are not stressed when an external magnetic field is applied (or conversely, that external stresses do not disrupt the magnetic properties). Our material

should find applications in miniaturization of electromechanical devices and in high-density magnetic data storage.

Electrochemical methods based on the early technology for fabricating magnetic wire memories^{15,16}, have long been used for practical production of conventional permalloy ($\text{Ni}_{80}\text{Fe}_{20}$ (at.%) films. Electroplated permalloy films have played a key role in the manufacture of magnetic recording heads and micro-magnetic actuators¹⁷. Many permalloy plating baths used industrially contain organic additives with sulphone functional groups, which reduce internal stress in the deposited film and thus decrease the grain size and lower the coercivity¹⁸. Such sulphur-containing additives (SCAs) are known to cause the inclusion of sulphur in the film through their surface activities^{18–21}. We have previously reported the magnetic properties of CoNiFe films prepared from baths containing saccharin or thiourea as SCAs^{22,23}. It was shown that the region of $B_s > 1.8$ T for CoNiFe and CoNiFeS films electroplated from saccharin and thiourea baths, respectively, is located in the Fe-rich and Co-rich areas. However, H_c values smaller than 160 A m^{-1} (2 Oe) could not be obtained. Our current success was achieved by formulating a new plating bath without SCAs and optimizing its operating conditions to produce films with fine crystal grains.

The composition and operating conditions of the basic bath for electroplating the CoNiFe film (see Methods) are based on those

described^{22,23} previously. Figure 1a shows a three-element map, indicating the regions with H_c smaller than 160 A m^{-1} for CoNiFe films electroplated from a bath containing saccharin (region I) and containing no SCA (region II). Region I is located in the low- B_s area (< 1.8 T), whereas region II is in the high- B_s area (> 2.0 T). The B_s value of the new film plated from SCA-free bath is greater than those reported for sputtered and electrodeposited soft magnetic films^{1–14}.

The phase boundary between face-centred cubic (f.c.c.) and body-centred cubic (b.c.c.) structures of a film plated from an SCA-free bath is shifted to the left compared with a film plated from the saccharin bath (Fig. 1b). The lines of zero saturation magnetostriiction, λ_s , are also shown in Fig. 1b; the removal of SCA shifts this line towards higher Fe content. These results show that a low- H_c CoNiFe film with B_s as high as 2.0–2.1 T and λ_s as low as 10^{-6} can be prepared without the use of SCA. One of the best films prepared has a composition of $\text{Co}_{65}\text{Ni}_{12}\text{Fe}_{23}$ with $B_s = 2.1$ T, $H_s = 95 \text{ A m}^{-1}$ (1.20 Oe), $\lambda_s = 1.8 \times 10^{-6}$, initial permeability $\mu_i = 600$ at 1 MHz, anisotropy field $H_k = 1,200 \text{ A m}^{-1}$ (15 Oe), and resistivity $\rho = 21 \mu\Omega \text{ cm}$. In addition, the film shows good thermal stability of magnetic properties up to 270°C . We will denote optimum films with these properties ‘HB-CoNiFe’ (high- B_s CoNiFe). Figure 2 shows a high-resolution transmission electron micrograph of an HB-CoNiFe film, composed of crystal grains with diameters of

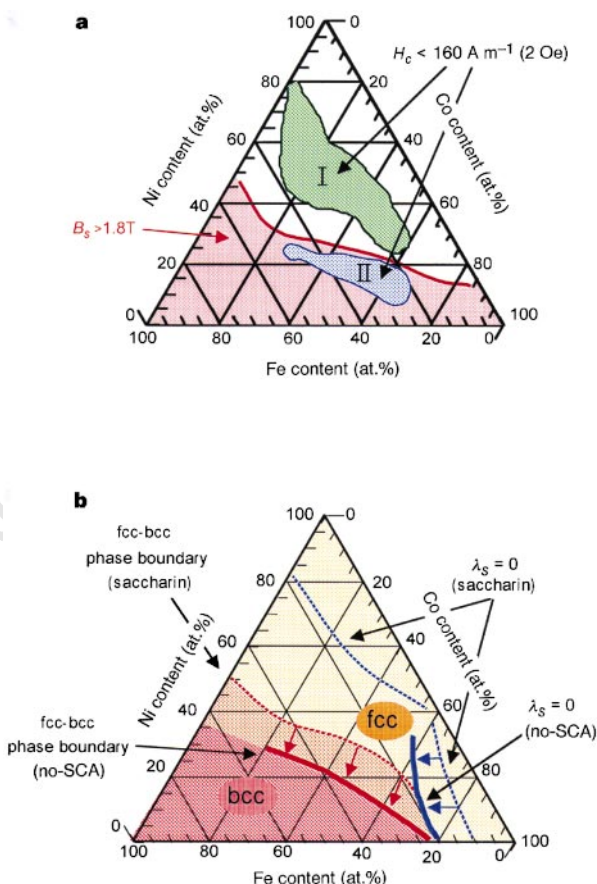


Figure 1 Three-element maps showing compositions of our CoNiFe films. **a**, Map showing low- H_c regions ($< 160 \text{ A m}^{-1}$) of two CoNiFe films produced by electroplating in baths with saccharin (region I) and with no SCA (region II). Region I is located within the $B_s = 1.0$ – 1.7 T region in the f.c.c. single-phase area. Region II is located in the high- B_s (> 2.0 T) region. **b**, Map showing f.c.c.–b.c.c. phase boundaries and lines of zero λ_s . The f.c.c.–b.c.c. phase boundary in the film produced from the no-SCA bath is shifted to the left, towards the Ni-poor region. The $\lambda_s = 0$ line of the film produced from the no-SCA bath is also shifted to the left, towards the Fe-rich region.

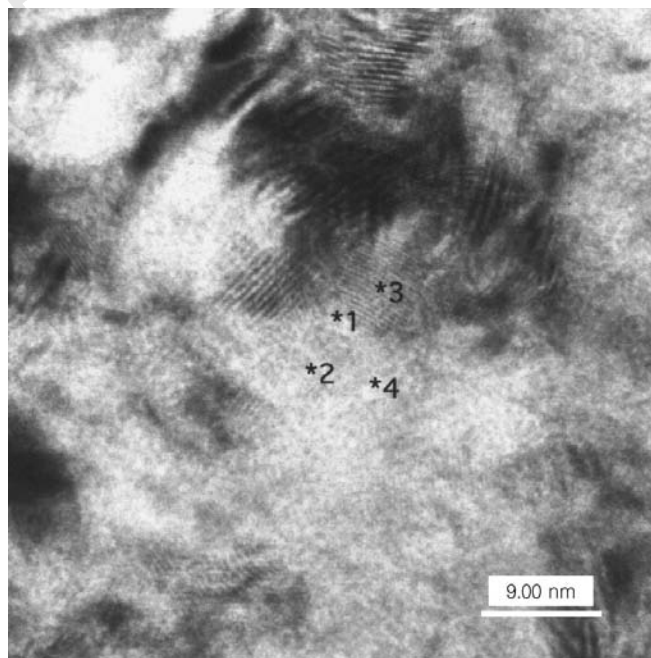


Figure 2 High-resolution transmission electron micrograph of the electroplated HB-CoNiFe film. A portion of the film consisting of fine crystal grains is seen at the centre of the image.

Table 1 Composition of HB-CoNiFe film

Location	Average composition (at.%)		
	Co	Ni	Fe
Centre of grain*	57.6	12.9	29.5
Boundary†	56.7	13.3	30.0

Compositions were determined by micro-EDS.

* Points 2 and 3 in Fig. 2.

† Points 1 and 4 in Fig. 2.

10–20 nm. The results of compositional analysis performed by EDS on spots measuring 1 nm in diameter are listed in Table 1. The areas for the analysis were selected at the centres of grains and at grain boundaries. The compositions were essentially identical at the two different locations, indicating that the CoNiFe film is highly homogeneous in composition.

We also investigated the corrosion resistance of this film in a 2.5% NaCl solution. Figure 3 shows anodic polarization curves of the HB-CoNiFe film, the electroplated CoNiFe film prepared in the saccharin bath, and the permalloy film electroplated from the bath also containing saccharin. The conventional CoNiFe film had the same composition as the HB-CoNiFe film except for the inclusion of a small amount of sulphur in the former (CoNiFe, 0.3 at.% sulphur; permalloy, 0.1 at.% sulphur; HB-CoNiFe, <0.1 at.% sulphur). Comparison of the polarization curves for the two types of CoNiFe ternary alloy films show that the anodic dissolution potential of the HB-CoNiFe film, –870 mV versus the saturated calomel electrode, is slightly more cathodic than that of the permalloy and the CoNiFe films, –820 mV, whereas the pitting corrosion potential (indicated by short unlabelled arrows in Fig. 3) of the HB-CoNiFe film is –65 mV, which is much more anodic than that of the CoNiFe film (–420 mV). The pitting corrosion potential is more important in practical applications than the dissolution potential. The pitting corrosion potential of the HB-CoNiFe film is even more anodic than that of the permalloy film. The results show that the HB-CoNiFe film has the best corrosion resistance of the three films tested, probably as a result of the low (or zero) sulphur content.

These new films may meet the requirements for miniaturization of step motors, micromagnetic sensors, transformers and magnetic recording heads. We have constructed prototype magnetoresistive heads in which the write head core is fabricated from the electroplated HB-CoNiFe films. These heads show superior performance characteristics²⁴. The achievement of magnetic field strengths of 2.0–2.1 T in a soft magnetic film provides a step towards realizing the targeted storage density of greater than 40 Gbit inch^{–2} in magnetic recording by the beginning of the twenty-first century²⁵. □

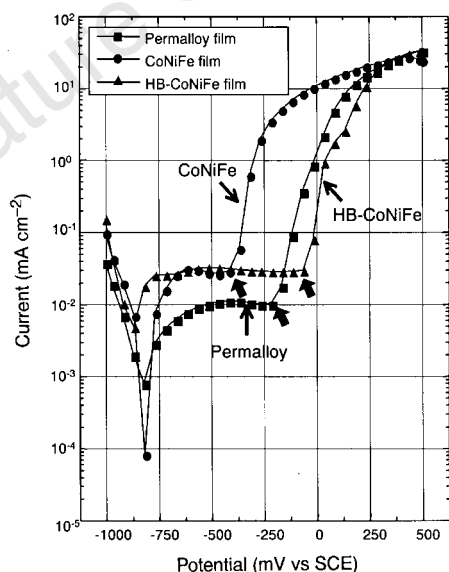


Figure 3 Anodic polarization studies. Shown are curves of electroplated permalloy, CoNiFe and HB-CoNiFe films in 2.5% NaCl solution, recorded by scanning the potential in the anodic direction at 10 mV s^{–1} at ambient temperature. The sequence of anodic dissolution potentials is HB-CoNiFe (–870 mV) > permalloy (–820 mV) ≈ CoNiFe (–820 mV), whereas that of pitting corrosion potentials is HB-CoNiFe (–65 mV) > permalloy (–170 mV) > CoNiFe (–420 mV).

Methods

The composition of the solution in the electroplating bath was as follows: cobalt sulphate 0.03–0.0875 mol dm^{–3}, nickel sulphate 0.2 mol dm^{–3}, ferrous sulphate 0.005–0.045 mol dm^{–3}, ammonium chloride 0.28 mol dm^{–3}, boric acid 0.4 mol dm^{–3}, sodium lauryl sulphate 0.01 g dm^{–3}. The current density and bath pH were varied from 3 to 20 mA cm^{–2} and from 2.5 to 3.0, respectively, and the temperature was 20 °C. The bath was operated with agitation using a rotating-disk electrode^{22,23} and a paddle plating cell²⁶. Film thickness was varied from 0.3 to 3.0 μm with a typical thickness of 1 μm. Values of B_s and H_c were measured by a vibrating sample magnetometer with an Earth magnetic field compensation circuit. Measurement reproducibility of the values of B_s and H_c were ±0.05 T and ±4 A m^{–1}, respectively. Values of λ_s were determined by an optical cantilever method in a 100 Oe rotating in-plane field. The reproducibility of λ_s was ±0.2 × 10^{–6}. The initial permeability was measured by the 8-figure coil method²⁷, which had a measurement reproducibility of ±5%. The specific resistance was measured by a four-probe method. The compositional analysis of Co, Ni and Fe was performed mainly by energy dispersive spectroscopy (EDS), and sulphur was determined by combustion analysis.

Received 6 October 1997; accepted 11 February 1998.

- Liao, S. H. High moment CoFe thin films by electrodeposition. *IEEE Trans. Magn.* **MAG-23**, 2981–2983 (1987).
- Chang, J. W., Andricacos, P. C., Petek, B. & Romankiw, L. T. in *Proc. Symp. on Magnetic Materials, Processes and Devices* (eds Romankiw, L. T. & Herman, D. A. Jr) 361–372 (PV90-8, Electrochem. Soc., Pennington, NJ, 1990).
- Shinoura, O. & Kamijima, A. Magnetic properties of electrodeposited CoFe film. *J. Surf. Finishing Soc. Jpn* **44**, 1114–1118 (1993).
- Chang, J. W., Andricacos, P. C., Petek, B. & Romankiw, L. T. in *Proc. 2nd Int. Symp. on Magnetic Materials, Processes, and Devices* (eds Romankiw, L. T. & Herman, D. A. Jr) 275–287 (PV92-10, Electrochem. Soc., Pennington, NJ, 1992).
- Anderson, N. C. & Chesnutt, R. B. Electrodeposition bath for cobalt-nickel-iron alloys for thin-film heads. US Patent no. 4,661, 216 (1987).
- Omata, Y. *Magnetic Properties of Fe-Co-Ni Films with High Saturation Magnetization Prepared by Evaporation and Electrodeposition* 23–29 (Tech. Rep. MR-23, 23–29, (1988).
- Shinoura, O., Kamijima, A. & Narumiya, Y. Soft magnetic properties of electrodeposited CoNiFe films. *J. Magn. Soc. Jpn* **18**, 277–280 (1994).
- Yoshizawa, Y., Oguma, S. & Yamauchi, K. New Fe-based soft magnetic alloys composed of ultrafine grain structure. *J. Appl. Phys.* **64**, 6044–6046 (1988).
- Ishiwata, N., Wakabayashi, C. & Urai, H. Soft magnetism of high-nitrogen-concentration FeTaN films. *J. Appl. Phys.* **69**, 5616–5618 (1991).
- Hasegawa, N. & Saito, M. Soft magnetic properties of microcrystalline Fe-M-C (M = V, Nb, Ta) films with high thermal stability. *J. Magn. Soc. Jpn* **14**, 313–318 (1990).
- Makino, A., Suzuki, K., Inoue, A. & Masumoto, T. Soft magnetic properties of bcc Fe-Zr-B sputtered films with nanoscale grain size. *Mater. Trans. Jpn Inst. Metals* **33**, 80–86 (1992).
- Takahashi, M., Shoji, H., Abe, M., Komaba, H. & Wakiyama, T. Soft magnetic preparation of Fe-N thin films sputtered in Ar + N₂ plasma. *J. Magn. Soc. Jpn* **14**, 283–288 (1990).
- Wang, S. & Kryder, M. H. RF-diode-sputtered iron nitride films for thin-film recording head materials. *J. Appl. Phys.* **67**, 5134–5136 (1990).
- Ogura, T., Kume, M. & Kuroki, K. Soft magnetic properties and domain structure of Fe-N-O thin films for thin film head applications. *J. Magn. Soc. Jpn* **18**(S1), 195–198 (1994).
- Wolf, I. W. Electrodeposition of magnetic materials. *J. Appl. Phys.* **33**, 1152–1159 (1962).
- Long, T. R. Electrodeposited memory element for a nondestructive memory. *J. Appl. Phys.* **31**(suppl.), 123–124 (1960).
- Romankiw, L. T. Electroformation of electronic devices. *Plating Surf. Finish.* **84**, 10–16 (1997).
- Smith, R. S., Godyski, L. E. & Lloyd, J. C. Effect of saccharin on the structural and magnetic properties of iron-nickel films. *J. Electrochem. Soc.* **108**, 996–998 (1961).
- Edwards, J. Radiotracer study of addition agent behaviour: 2-adsorption on metal surface from solutions containing thiourea. *Trans. Inst. Met. Finish.* **39**, 45–51 (1962).
- Edwards, J. Radiotracer study of addition agent behaviour: 3-incorporation of sulphur in nickel deposited from solutions containing p-toluenesulphonamide and saccharin. *Trans. Inst. Met. Finish.* **39**, 52–55 (1962).
- Brown, H. Addition agents, anions, and inclusions in bright nickel plating. *Plating* **55**, 1047–1055 (1968).
- Nakamura, A., Takai, M., Hayashi, K. & Osaka, T. Preparation and magnetic properties of CoNiFe thin film by electrodeposition. *J. Surf. Finish. Soc. Jpn* **47**, 934–938 (1996).
- Takai, M., Hayashi, K., Aoyagi, M. & Osaka, T. Electrochemical preparation of soft magnetic CoNiFeS film with high saturation magnetic flux density and high resistivity. *J. Electrochem. Soc.* **144**, L203–L204 (1997).
- Ohashi, K. et al. Newly developed CoNiFe film for MR head. Paper no. DQ-09, presented at the 7th Joint MMM-Intermag Conf., San Francisco, January 1998.
- Roundtable Discussion about the Future of Magnetic Storage, Session ZA-02, Presented at the 7th Joint MMM-Intermag Conf., San Francisco, January 1998.
- Powers, J. V. & Romankiw, L. T. *Electroplating Cell Including Means to Agitate the Electrolyte in Laminar Flow* US Patent no. 4,652,442 (1972).
- Calcagno, P. A. & Thompson, D. A. Semiautomatic permeance tester for thick magnetic films. *Rev. Sci. Instrum.* **46**, 904–908 (1975).

Acknowledgements. We thank Y. Okinaka, Waseda University, for discussions. This work was supported by a Research for Future Project of the Japan Society for Promotion of Science, a Grant-in-Aid for Scientific Research on Priority Area from the Ministry of Education, Science, Sports and Culture, Japan, and the Storage Research Consortium (SRC).

Correspondence and requests for materials should be addressed to T.O. (e-mail: osakatet@mn.waseda.ac.jp).

Biomimetic engineering of non-adhesive glycocalyx-like surfaces using oligosaccharide surfactant polymers

Nolan B. Holland*, Yongxing Qiu†, Mark Ruegsegger† & Roger E. Marchant*†

Departments of *Macromolecular Science and †Biomedical Engineering, Case Western Reserve University, Cleveland, Ohio 44106, USA

The external region of a cell membrane, known as the glycocalyx, is dominated by glycosylated molecules^{1–3}, which direct specific interactions such as cell–cell recognition and contribute to the steric repulsion that prevents undesirable non-specific adhesion of other molecules and cells. Mimicking the non-adhesive properties of a glycocalyx provides a potential solution to the clinical problems, such as thrombosis⁴, that are associated with implantable devices owing to non-specific adsorption of plasma proteins. Here we describe a biomimetic surface modification of graphite using oligosaccharide surfactant polymers, which, like a glycocalyx, provides a dense and confluent layer of oligosaccharides. The surfactant polymers consist of a flexible poly(vinyl amine) with dextran and alkanoyl side chains. We show that alkanoyl side chains assemble on graphite through hydrophobic interaction and epitaxial adsorption. This constrains the polymer backbone to lie parallel to the substrate, with solvated dextran side chains protruding into the aqueous phase, creating a glycocalyx-like coating. The resulting biomimetic surface is effective in suppressing protein adsorption from human plasma protein solution.

A significant challenge in engineering a glycocalyx-like surface on biomaterials used for implantation is to achieve a high surface density of oligosaccharides. Typical biomaterials are hydrophobic, with no labile chemical groups that might be used with conventional immobilization strategies⁵. The design of our surfactant polymers (Fig. 1) includes a flexible poly(vinyl amine) backbone with alkanoyl side chains that serve as physical binding ligands to a hydrophobic substrate, and oligosaccharide side chains that protrude into the aqueous environment to provide a glycocalyx-like interface. The surface modification construct is demonstrated here with dextran as the oligosaccharide and highly orientated pyrolytic graphite as the substrate, but the same approach may be

extended to include different oligosaccharides and other hydrophobic surfaces.

The surfactant polymers, poly(*N*-vinyl dextran aldonaide-co-*N*-vinyl alkanamide) shown in Fig. 1c, are synthesized using well defined poly(vinyl amine) (ref. 6; number-average molecular mass $M_n \approx 6,000$ daltons) followed by attaching oligosaccharides of dextran ($9\alpha(1 \rightarrow 6)$ glucose residues; mass-average molecular mass, 1,600 daltons; polydispersity = 1.16) and alkanoyl (hexanoyl or lauroyl) chains simultaneously and selectively to reactive amine sites⁷. Selective attachment of dextran is accomplished by reacting amine groups with dextran lactone, whereas alkanoyl groups are attached using *N*-alkanoyloxy succinimide. We varied the molar feed ratios to produce different ratios of side chains on the polymer backbone. In this manner, five surfactant polymers were prepared with different side chain ratios of dextran to hexanoyl (dex-hex) and dextran to lauroyl (dex-lau).

Surfactant polymers were characterized by Fourier-transform infrared (FTIR) and ¹H-NMR spectroscopies and elemental analysis to confirm purity and structure. The surfactant polymers are surface active at an air–water interface, as determined by significant reductions in water surface tension, and undergo spontaneous adsorption on diverse hydrophobic surfaces such as highly orientated pyrolytic graphite and low-density polyethylene (PE). Pyrolytic carbon and PE are used in implantable prosthetic devices. The hydrophilic nature of a modified surface was confirmed by measurement of low water contact angles. The stability on PE was demonstrated by the absence of detectable desorption under conditions of pure water flowing across the modified surface⁸.

Real-time images of the surfactant polymers adsorbing to highly orientated pyrolytic graphite were obtained using an atomic force microscope (AFM; Nanoscope III, Digital Instruments) operated in fluid tapping mode⁹. Freshly cleaved graphite was imaged under water before the addition of surfactant solution. Imaging of adsorbed surfactant began after the microscope had equilibrated with the new solution (~10 min). The results for the surfactant polymer with 1:5 dex–hex adsorbed on graphite from a 0.5 mg ml^{–1} aqueous solution are shown in Fig. 2. The initial scans (Fig. 2a), obtained after ~25 min, show strands of surfactant polymer adsorbed on the surface. As time progressed, the strands broaden (Fig. 2b, c), and by ~7 h (Fig. 2d) most of the surface is covered. Residual surface roughness disappears by 20 h, leaving a complete and compact monolayer. The thickness of the adsorbed surfactant polymer increases slightly with increasing surface coverage up to a maximum of 7–12 Å. This indicates a single monolayer on the surface, consistent with our estimate for the dextran side-chain diameter of ~9 Å, obtained from surface-tension data of *N*-dodecyl dextran aldonaide¹⁰, and ~2.5 Å for the polymer backbone. Evidence that the adsorption is stable was obtained

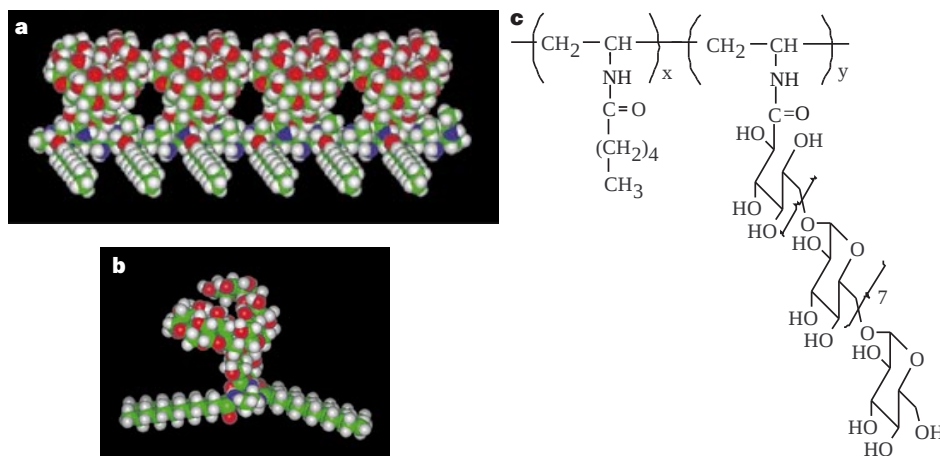


Figure 1 Molecular models of oligosaccharide surfactant polymer. This polymer consists of a poly(vinyl amine) backbone with multiple dextran and alkanoyl side chains, randomly spaced. The molecular models show a portion of a 1:1.5 dex–lau surfactant polymer (surface-adsorbed conformation) as a side view (a) and an end view (b), with the chemical composition shown in c. Molecular models were generated using BIOSYM software. (Colour code: carbon, green; oxygen, red; nitrogen, blue; and hydrogen, white).

from observing no discernible change in the monolayer after keeping pure water over the adsorbed surfactant for 1–2 h and scanning with high AFM imaging forces.

The ordering of adsorbing surfactant polymer on graphite was unexpected. Polymers with relatively short-chain branches do not tend to crystallize, because branches prevent close intermolecular alignment of the polymer backbone. On the graphite, however, we observe an adsorption pattern of the polymer in strands 60° out of alignment, verified by the two-dimensional Fourier transform of the image showing hexagonal angular dependence (Fig. 2c, inset). Images of the graphite lattice verify that the strands align perpendicular to the substrate atoms. This is consistent with a mechanism

in which alkanoyl side chains adsorb in registry with the graphite atoms, which constrains the polymer backbone to the surface and orientates dextran side chains away from the surface (Fig. 3). A large enthalpic driving force (6.28 kJ mol⁻¹ per methylene group)¹¹ for epitaxial ordering of hydrocarbons is supported by many reported examples^{11–17}, including paraffin chain alignment on graphite^{11–13} and epitaxial crystallization of linear polyethylene^{14,15}.

Surfactant polymer ordering on graphite propagates laterally, with preferential adsorption at the edge of previously adsorbed polymer. Epitaxial growth patterns were observed for three surfactant polymers that contained a relatively high alkanoyl chain concentration (1:5 and 1:4 dex–hex and 1:1.5 dex–lau). Surfactant

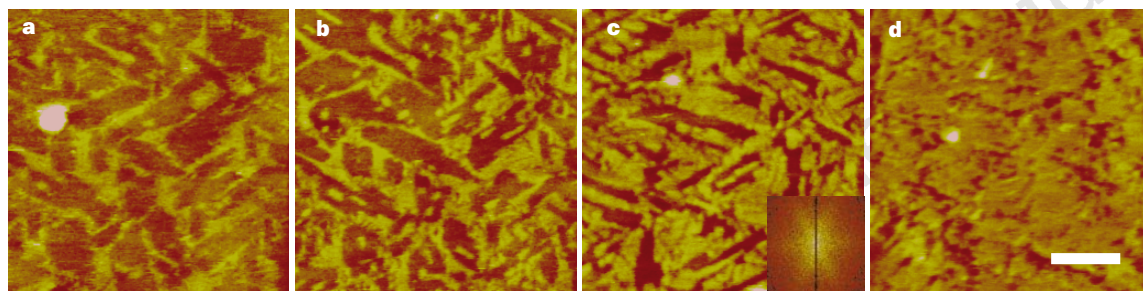


Figure 2 AFM images, observed *in situ*, of adsorbed surfactant polymer. This polymer is poly(*N*-vinyl dextran aldonamide-co-*N*-vinyl hexanamide) with 1:5 dextran to hexanoyl side-chain ratio (1:5 dex–hex), adsorbed on the basal plane of graphite. The images (scale bar, 250 nm) show the progression of adsorption from: **a**, ~25 min; **b**, ~1 h; **c**, ~5.5 h; and **d**, ~7.5 h; as imaged in 0.5 mg ml⁻¹ polymer solution. **a**, Surfactant polymer molecules initially adsorb in strands, driven by hydrophobic interaction and epitaxial adsorption of hexanoyl side chains onto the

graphite. **b, c**, Additional polymer adsorbs along previously adsorbed chains, as observed by the increasing widths of the adsorbate strips. Epitaxial adsorption is verified by the hexagonal pattern seen in the two-dimensional Fourier transform (**c**, inset). **d**, The graphite surface is almost fully covered by surfactant polymer. Complete monolayer coverage was observed after 20 h. The height of the adsorbed surfactant polymer is 7–12 Å, indicating a single monolayer.

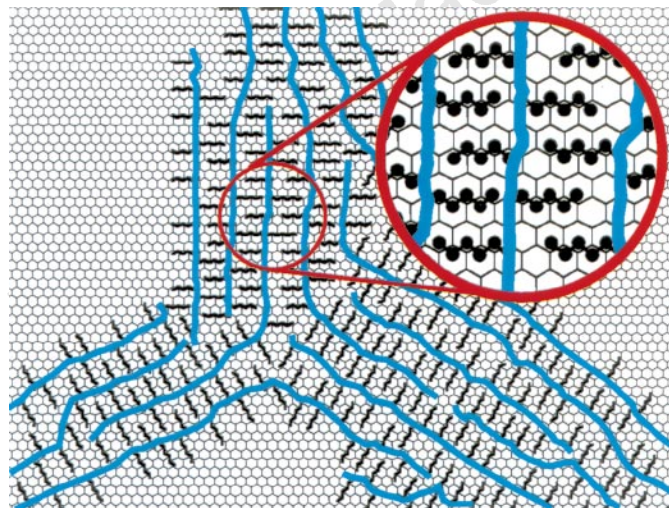


Figure 3 Model of the epitaxial adsorption of surfactant polymer on graphite. The extension and alignment of the poly(vinyl amine) backbone (thick lines) are driven by the preferential adsorption of hexanoyl side chains in registry with the atomic lattice of the graphite (see expanded view). The relatively high density of randomly spaced hydrocarbon side chains requires the polymer backbone to extend to allow each hexanoyl to be in registry. The entropy lost in extending the polymer chain is offset by the enthalpic energy gained from epitaxial adsorption. Hexanoyl side chains can align in three symmetric directions on the graphite lattice, leading to the alignment of the polymer backbone in three directions as observed by AFM. Dextran side chains attached to the backbone (not shown) are oriented away from the substrate into the aqueous solution.

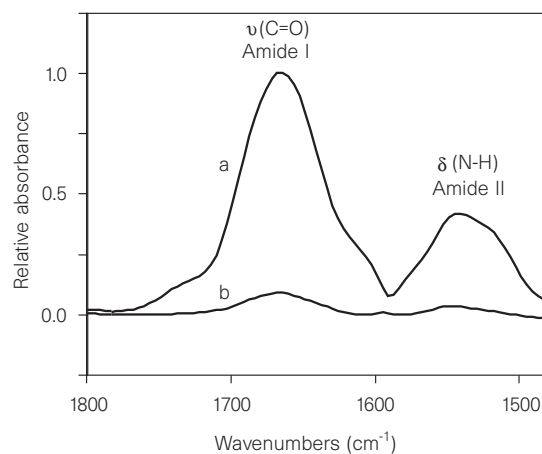


Figure 4 Infrared spectra (1800–1400 cm⁻¹ region) showing amount of protein adsorption. Trace a, spectrum showing high adsorption of plasma proteins from 50% fresh human platelet-poor plasma on graphite, and trace b spectrum showing negligible protein adsorption on graphite surface modified with surfactant polymer (1:5 dex–hex). Protein adsorption was determined from the relative absorbance intensity of the protein amide I (1650 cm⁻¹) and amide II (1550 cm⁻¹) bands on each surface. Infrared spectra (ATR mode) were collected by co-adding 100 scans with a resolution of 8 cm⁻¹, using an FTS 575C infrared spectrometer, equipped with a UMA 500 microscope (Bio-Rad). Spectra were normalized using the sharp negative peak (1590 cm⁻¹) characteristic of the graphite²² as an internal standard, followed by spectral subtraction of water vapour and the graphite substrate to reveal the protein adsorbate. For trace b, the spectral contribution from amide bonds in the surfactant polymer was also subtracted.

polymers with relatively low alkanoyl concentration (1:1 dex–hex and 1:0.6 dex–lau) adsorb to the surface, but not in an ordered pattern. In this case, alkanoyl chains may bind in registry to achieve energy minimization, but are spaced too far apart to induce the polymer backbone to extend. To achieve ordering, enthalpic energy gained from epitaxial adsorption of alkanoyl side chains offsets the entropy loss in extending the polymer backbone and forcing dextran side chains into close proximity. Steric repulsion between dextrans may result in a 'brush-like' conformation, although there is little comparison with the extended conformations shown by surface-anchored synthetic random coil polymers interacting with good solvents^{18–20}. Dextran has a stable helical structure that is unlikely to be perturbed by nearest-neighbour effects.

A design feature of the biomimetic surface is to use the protruding dextran oligosaccharides to mimic the non-adhesive properties of a glycocalyx. The steric barrier provided by the highly hydrated dextrans is designed to suppress non-specific adsorption of plasma proteins²¹, whereas the high energy of desorption and low water solubility of the adsorbed surfactant polymer is designed to minimize possible displacement or exchange reactions with highly surface-active plasma proteins. We tested this biomimetic concept for graphite modified with a confluent layer of 1:5 dex–hex obtained in a 24-h adsorption from aqueous solution. Using a laminar flow cell, a 50% solution of fresh human platelet-poor plasma in phosphate-buffered saline (pH 7.4) anticoagulated with sodium citrate, was adsorbed on graphite samples under static conditions for 30 min at 37 °C. This provided a rigorous *in vitro* test, as the solution contained all plasma proteins in blood at a concentration that was sufficient to cover a surface with proteins every second of exposure. The protein solution was replaced by phosphate-buffered saline and the samples were rinsed under a controlled shear stress of 10 dyne cm⁻² for 5 min. Samples were air-dried overnight and under partial vacuum.

Protein adsorption was quantified from the relative infrared (attenuated total internal reflection mode) absorbance intensity of the characteristic protein 'amide I' (1650 cm⁻¹) and 'amide II' (1550 cm⁻¹) bands. To reveal the protein adsorbate (Fig. 4), spectra were normalized to a sharp negative peak (1590 cm⁻¹) characteristic of the graphite²², followed by digital subtraction of water vapour and the graphite substrate. On bare graphite, the amide I and II bands are attributed entirely to adsorbed proteins, whereas on modified graphite amide bonds in the surfactant polymer and adsorbed proteins contribute to the absorbance. By subtracting the contribution from the surface polymer, the absorbance due to adsorbed protein is isolated and determined. As shown in Fig. 4, amide I and II protein bands are very strong on bare graphite and almost negligible on the biomimetic surfactant-modified graphite. We estimate that plasma protein adsorption is suppressed by at least 90% on the biomimetic surface, compared with the bare graphite, which is consistent with a stably adsorbed surfactant polymer.

The glycocalyx-like modification of graphite demonstrates effective suppression of non-specific protein adsorption, although testing under *in vivo* conditions is needed to confirm the usefulness of this approach on implantable devices. The biomimetic approach offers many possibilities for manipulating the surfactant polymer design. The oligosaccharide composition of side chains could be varied to engineer a range of modified materials, each designed to mimic desirable properties of a glycocalyx. It should be possible to develop biomimetic designs that incorporate biologically important oligosaccharide sequences found in a particular cell-surface glycocalyx. One can also envisage surfactant polymer designs in which adhesive, rather than non-adhesive, biological interactions are desired, such as incorporating the anticoagulant pentasaccharide sequence of heparin, specific cell-adhesion receptors (such as selectins) or adhesive peptide sequences that could be attached to the polymer backbone through a polar appendage. □

Received 17 November 1997; accepted 16 February 1998.

1. Evans, W. H. & Graham, J. M. *Membrane Structure and Function* 1–86 (Oxford Univ. Press, New York, 1991).
2. Fukuda, M. in *Molecular Glycobiology* (eds Fukuda, M. & Hindsgaul, O.) 1–52 (Oxford Univ. Press, New York, 1994).
3. Bongrand, P. in *Physical Basis of Cell-Cell Adhesion* (ed. Bongrand, P.) 1–37 (CRC, Boca Raton, FL, 1987).
4. Salzman, E. W., Merrill, E. W. & Hent, K. C. in *Hemostasis and Thrombosis* 3rd edn (eds Colman, R. W., Hirsh, J., Marder, V. J. & Salzman, E. W.) 1469–1485 (Lippincott, Philadelphia, 1994).
5. Helmus, M. N. & Hubbell, J. A. Materials selection. *Cardiovasc. Pathol. (Suppl.)* **2**, 53S–71S (1993).
6. Badesso, R. & Pinschmidt Jr., R. K. Synthesis of amine functional homopolymers with N-ethenylformamide. *Polym. Mater. Sci. Eng.* **69**, 251–252 (1993).
7. Qiu, Y., Zhang, T., Ruegsegger, M. & Marchant, R. E. Novel nonionic oligosaccharide surfactant polymers derived from poly(vinyl amine) with pendant dextran and hexanoyl groups. *Macromolecules* **31**, 165–171 (1998).
8. Qiu, Y. & Marchant, R. E. Protein resistant biomaterial surfaces derived from physically immobilized novel oligosaccharide surfactant molecules. *Trans. Soc. Biomater.* **20**, 224 (1997).
9. Putman, C. A. J., Vanderwerf, K. O., De Grooth, B. G., Van Hulst, N. E. & Greve, J. Tapping mode atomic-force microscopy in liquid. *Appl. Phys. Lett.* **64**, 2454–2456 (1994).
10. Zhang, T. & Marchant, R. E. Novel polysaccharide surfactants: The effect of hydrophobic and hydrophilic chain length on surface active properties. *J. Colloid Interface Sci.* **177**, 419–426 (1996).
11. Groszek, A. Selective adsorption at graphite/hydrocarbon interfaces. *Proc. R. Soc. Lond. A* **314**, 473–498 (1970).
12. McGonigal, G. C., Bernhardt, R. H. & Thompson, D. J. Imaging alkane layers at the liquid/graphite interface with the scanning tunneling microscope. *Appl. Phys. Lett.* **57**, 28–30 (1990).
13. Rabe, J. P. & Buchholz, S. Commensurability and mobility in two-dimensional molecular patterns on graphite. *Science* **253**, 424–427 (1991).
14. Wellington, S., Rybníkar, F. & Baer, E. Epitaxial crystallization of polyethylene. *J. Macromol. Sci. Phys. B* **10**, 1–39 (1974).
15. Whitman, J. C. & Lotz, B. Epitaxial crystallization of polymers on organic and polymeric substrates. *Prog. Polym. Sci.* **15**, 909–948 (1990).
16. Manne, S. & Gaub, H. E. Molecular organization of surfactants at solid-liquid interfaces. *Science* **270**, 1480–1482 (1995).
17. Manne, S., Cleveland, J. P., Gaub, H. E., Stucky, G. D. & Hansma, P. K. Direct visualization of surfactant hemimicelles by force microscopy of the electrical double-layer. *Langmuir* **10**, 4409–4413 (1994).
18. Milner, S. T. Polymer brushes. *Science* **251**, 905–914 (1991).
19. Sharma, R. Small-molecule surfactant adsorption, polymer surfactant adsorption, and surface solubilization — An overview. *ACS Symp. Ser.* **615**, 1–20 (1995).
20. Golander, C.-G. *et al.* in *Poly(ethylene glycol) Chemistry* (ed. Harris, J. M.) 221–246 (Plenum, New York, 1992).
21. Österberg, E. *et al.* Protein-rejecting ability of surface bound dextran in end-on and side-on configurations: Comparison to PEG. *J. Biomed. Mater. Res.* **29**, 741–747 (1995).
22. Nemanich, R. J., Lucovsky, G. & Solin, S. A. Infrared active optical vibrations of graphite. *Solid State Commun.* **23**, 117–120 (1977).

Acknowledgements. We acknowledge the financial support for this work provided by the National Institutes of Health and the Whitaker Foundation, and use of the AFM and IR facilities at the Center for Cardiovascular Biomaterials, CWRU.

Correspondence and requests for materials should be addressed to R.E.M. (e-mail: rxm4@po.cwru.edu).

Direct linking of microbial populations to specific biogeochemical processes by ¹³C-labelling of biomarkers

H. T. S. Boschker^{*†}, S. C. Nold^{*}, P. Wellsbury[‡], D. Bos^{*}, W. de Graaf^{*}, R. Pel^{*}, R. J. Parkes[‡] & T. E. Cappenberg[†]

^{*} Centre for Limnology, Netherlands Institute of Ecology, Rijksstraatweg 6, 3631 AC Nieuwersluis, The Netherlands

[†] Centre for Estuarine and Coastal Ecology, Netherlands Institute of Ecology, Postbus 140, 4400 AC Yerseke, The Netherlands

[‡] Department of Geology, University of Bristol, Bristol BS8 1RJ, UK

Recent advances in the application of molecular genetic approaches have emphasized our potentially huge underestimate of microbial diversity in a range of natural environments¹. These approaches, however, give no direct information about the biogeochemical processes in which microorganisms are active². Here we describe an approach to directly link specific environmental microbial processes with the organisms involved, based on the stable-carbon-isotope labelling of individual lipid biomarkers. We demonstrate this approach in aquatic sediments and provide evidence for the identity of the bacteria involved in two important biogeochemical processes: sulphate reduction coupled to acetate

oxidation in estuarine and brackish sediments^{3,4}, and methane oxidation in a freshwater sediment⁵. Our results suggest that acetate added in a ¹³C-labelled form was predominantly consumed by sulphate-reducing bacteria similar to the Gram-positive *Desulfotomaculum acetoxidans* and not by a population of the more widely studied Gram-negative *Desulfobacter* spp. Furthermore, ¹³C-methane labelling experiments suggest that type I methanotrophic bacteria dominate methane oxidation at the freshwater site.

In the first application, sulphate-reducing bacteria were targeted to demonstrate this approach because they have a range of important characteristics. First, the lipids of these bacteria, especially the polar lipid derived fatty acids (PLFAs), have been studied extensively and several 17-carbon compounds have been proposed as specific biomarkers for genera in the large Gram-negative group^{6–8}. Second, sulphate reduction is an environmentally significant process, as sulphate-reducing bacteria can be responsible for ~50% of the total carbon mineralization in more active coastal sediments³. Third, sulphate-reducing bacteria in sediments use only a limited number of simple organic substrates, predominantly acetate, which are produced during the first fermentative steps in the anaerobic decomposition of organic matter^{4,9}.

At our intertidal estuarine site (Southdown, Tamar estuary, UK; salinity 35‰), anaerobic carbon mineralization in the sediment was dominated by sulphate reduction, which accounted for 10–60% (seasonal variation) of the total carbon mineralization as measured by sediment oxygen uptake (data not shown). Acetate oxidation and sulphate-reduction rates were similar in the 1–3 cm sediment layer, which suggested that acetate was the dominant substrate for the sulphate-reducing bacteria (data not shown; see Methods¹⁰). This was confirmed by the addition of 20 mM molybdate, a specific inhibitor of sulphate reduction¹¹, which completely inhibited acetate uptake. Moreover, acetate production was not inhibited by molybdate and could explain 92% of the original sulphate-reduction rate.

The pattern of PLFA concentrations in Tamar sediment (Fig. 1a) was similar to what is commonly found in marine sediments¹², and natural $\delta^{13}\text{C}$ ratios for the individual PLFAs were on average $-24.0 \pm 1.5\text{‰}$. After incubation of sediments with 100 μM ¹³C-labelled acetate, the $\delta^{13}\text{C}$ ratios in some PLFAs increased by up to 90‰ (limit of detectable enrichment, 1–2‰). The individual PLFAs that incorporated the most label were 16:1 ω 7c, 16:1 ω 5c, 16:0 and 18:1 ω 7c (Fig. 1b, c). Total labelling patterns in March were similar to those in August (Fig. 1b, c), except that 16:1 ω 5c and 18:1 ω 7c received comparatively less label. Labelling patterns in incubations receiving high ¹³C-acetate concentrations (100 μM) were similar to those incubated with 10 μM ¹³C-acetate (Fig. 2), which is in the range of the *in situ* porewater concentration of acetate (8–15 μM ; ref. 10). Labelling patterns with 100 μM ¹³C-acetate of sediments from a brackish lagoon at the island of Rügen, Germany were similar to Tamar sediments (Fig. 2), which suggests that similar populations of acetate-consuming sulphate reducers were active in both sediments. Surprisingly, none of the 17-carbon PLFAs, characteristic for many of the known Gram-negative sulphate reducers, were highly labelled, casting doubt on the importance of this bacterial group in marine sediments.

As the PLFAs that received most of the label are common in bacteria¹³, we tested if acetate incorporation was indeed coupled to sulphate reduction in Tamar sediments. Molybdate inhibited labelling of the four main labelled PLFAs by 55–100% in August (Fig. 1c, total inhibition over all PLFAs $60 \pm 20\%$), which confirmed that most label was incorporated by sulphate reducers. However, some incorporation by other anaerobic bacteria occurred, which might have been expected because acetate can be used by many organisms as an intermediate in lipid synthesis¹³. The part of the total acetate incorporation sensitive to inhibition by molybdate can be assigned most directly to sulphate-reducing populations and showed a PLFA

labelling pattern (grey bars, Fig. 1c) that was not very different from the total incorporation (complete bars, Fig. 1c; see also Fig. 2).

Our data suggest that acetate labelling patterns could be interpreted as the PLFA pattern of the dominant sulphate-reducing populations in Tamar and Rügen sediments. Labelling patterns were therefore compared with PLFA compositions of known sulphate-reducing bacteria (Fig. 2). Surprisingly, all sediment samples clustered

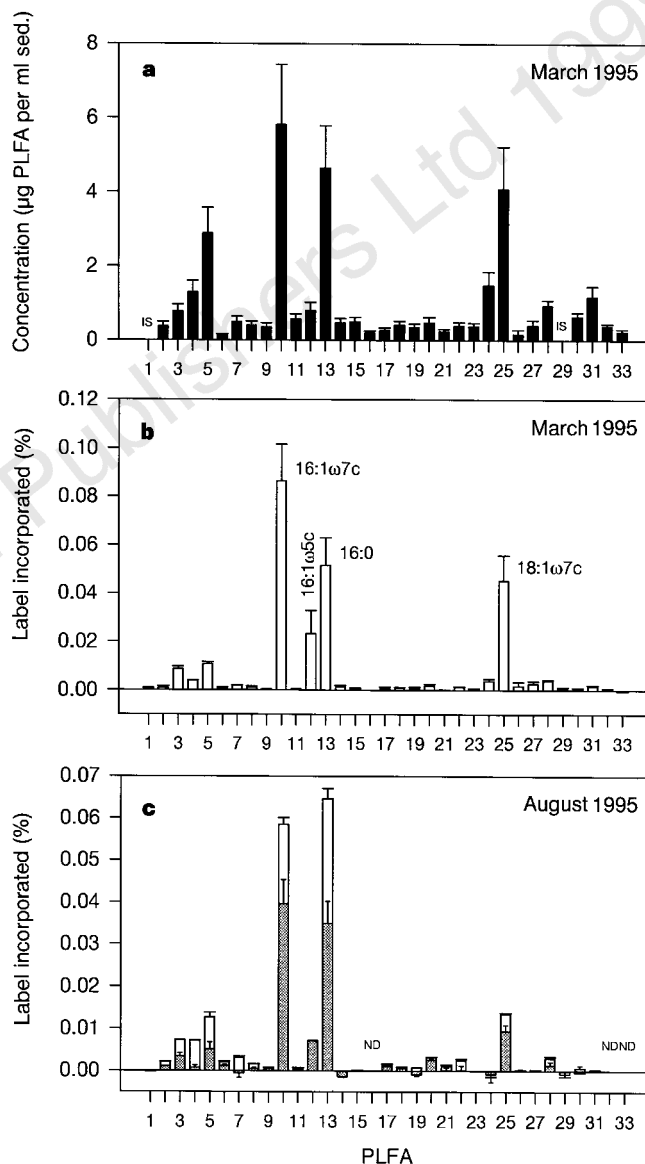


Figure 1 Results of the ¹³C-acetate labelling study in intertidal sediments of the Tamar Estuary, UK. **a**, PLFA concentrations for the 1–3 cm horizon of Tamar sediment, and **b**, percentage of added ¹³C-labelled acetate (100 μM , final concentration) incorporated into PLFAs in the same sediment layer, March 1995. **c**, Effect of molybdate, a specific inhibitor of sulphate reduction, on the incorporation of ¹³C-labelled acetate into PLFAs in the 1–3 cm horizon, August 1995. The grey part of the bars shows the incorporation sensitive to inhibition by molybdate, which is most likely of sulphate-reducing bacterial origin. This was calculated as the difference between label incorporation in the control (total height of bars) and molybdate incubations. Experimental details are given in the Methods section. Numbers on x-axis indicate PLFAs as follows (PLFA nomenclature as in ref. 12): 1, 12:0 (internal standard, IS); 2, 14:0; 3, 14:0; 4, 15:0; 5, 15:0; 6, 15:1 ω 8; 7, 15:0; 8, 16:0; 9, 16:1 ω 9c/5t; 10, 16:1 ω 7c/8c/6c (predominantly 7c); 11, 16:1 ω 7t; 12, 16:1 ω 5c; 13, 16:0; 14, 17:1 ω 7 + unknown; 15, 10Me16:0; 16, 17:1 ω 7; 17, 17:0; 18, 17:1 ω 8; 19, 17:1 ω 6; 20, 17:1 ω 6; 21, 17:0; 22, 17:0; 23, 18:2; 24, 18:1 ω 9c; 25, 18:1 ω 7c/8c (predominantly 7c); 26, 18:1; 27, 18:1; 28, 18:0; 29, 19:0 (internal standard); 30, 20:4 ω 6; 31, 20:5 ω 3; 32, 20:1; 33, 22:0 PUFA. ND, not detected.

with the Gram-positive *Desulfotomaculum acetoxidans*, which readily oxidizes and grows on acetate¹⁴ and which has almost exclusively the common PLFAs that incorporated most label⁷. The type strain of this species is generally not considered to be a sedimentary marine organism because of its growth characteristics¹⁵, but similar organisms have been isolated from marine sediments¹⁶ and we could easily enrich them on acetate from pasteurized Tamar sediments (data not shown). Acetate-using isolates from anaerobic marine sediments mostly belong to the Gram-negative genus *Desulfobacter*^{15,17,18}. However, ¹³C-acetate incorporation patterns clearly indicate that this genus was not the dominant acetate-consuming population in the two sediments studied (Fig. 2). We can not exclude the possibility that the labelling patterns were caused by the presence of an as-yet unknown genus of sulphate-reducing bacteria¹⁹. However, a recent study suggests that PLFA compositions are specific for the genera of sulphate-reducing bacteria⁸. It also seems unlikely that our labelling patterns were

the result of a mixture of many different genera, as this would have produced a more even label distribution among bacterial PLFAs. Large differences in growth yield between species would also affect our interpretation, but maximum growth yields are similar for *Desulfotomaculum acetoxidans* (5.6 g per mol acetate)¹⁴ and *Desulfobacter* spp. (4.8 g per mol acetate)¹⁵. In short, our results suggest that a *Desulfotomaculum acetoxidans*-like organism or group of organisms dominated acetate-coupled sulphate reduction in both Tamar and Rügen sediments. The Gram-positive group of sulphate-reducing bacteria to which this organism belongs has been neglected in recent sediment studies^{19–22}, but should be given more attention in future research.

In the second example of the approach, we studied the methane-oxidizing bacteria inhabiting the top layer of sediments in the eutrophic Lake Loosdrecht, The Netherlands. These microorganisms also have a number of characteristics that make them suitable for labelling studies. First, microbial methane oxidation is an

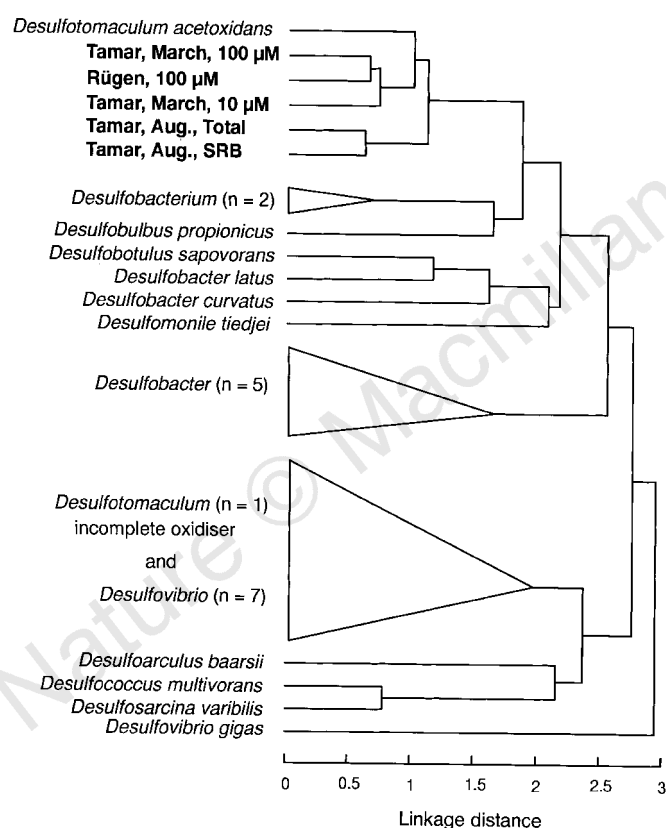


Figure 2 Comparison of ¹³C-acetate PLFA labelling patterns of Tamar sediments with published^{7,8} PLFA compositions of pure cultures of mesophilic sulphate-reducing bacteria. In addition, acetate labelling patterns are shown for sediment from a brackish lagoon on the island of Rügen, Germany (100 μM ¹³C-acetate, 1–3 cm depth, October 1994). The difference between the sediment/*Desulfotomaculum acetoxidans* cluster and the nearest, *Desulfobacterium*, cluster is based both on the presence of 17:1ω6 and on a lower amount of 18:1ω7c in the latter cluster. Only PLFAs between 12:0 and 19:0 were used. Cluster analysis was performed with the software package Statistica (StatSoft Inc.) on euclidean distances between log-transformed normalized data (mole-percentages) by the unweighted pair-group method using averages (UPGMA). The cluster containing the sediments and *Desulfotomaculum acetoxidans* was stable with respect to the clustering techniques used, and results leading to the same conclusions were obtained by multi-dimensional scaling analysis. The only Gram-positive genus, *Desulfotomaculum*, can be divided into two distinct metabolic groups: (1) complete oxidizing strains like *Desulfotomaculum acetoxidans* and (2) incomplete oxidizing, non-acetate-using strains. SRB, sulphate-reducing bacteria.

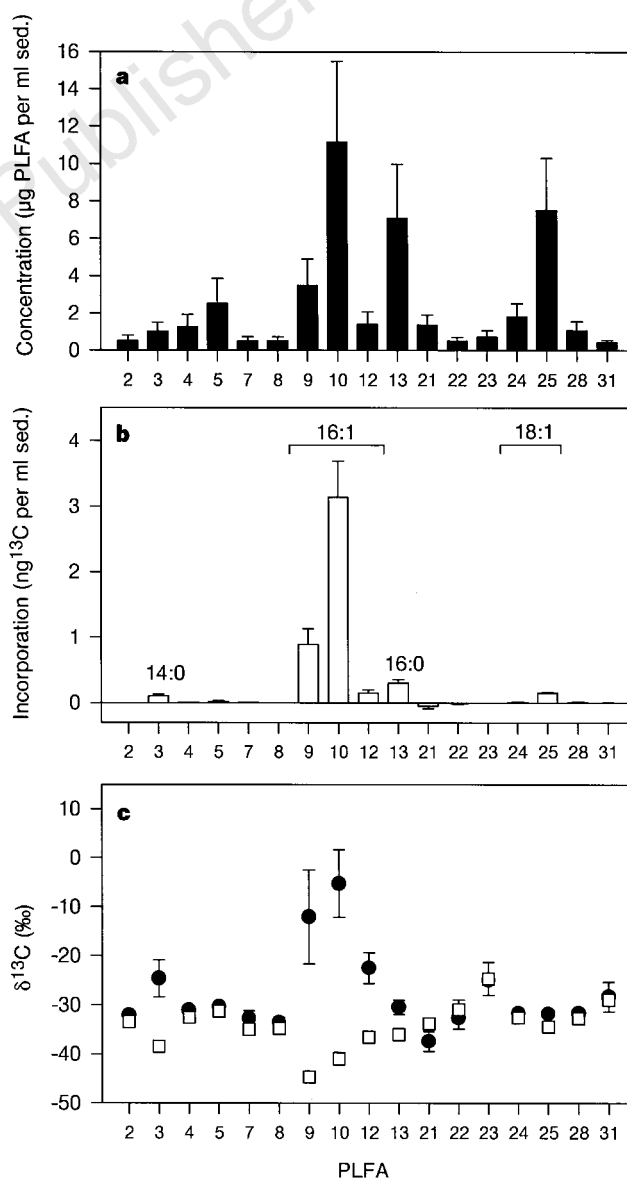


Figure 3 Results of the ¹³C-methane labelling study in Lake Loosdrecht sediment, The Netherlands. **a**, PLFA concentrations in the 0–0.5 cm horizon of the sediment, and **b**, amounts of ¹³C incorporated into the different PLFAs. **c**, Comparison between carbon-isotope ratios of PLFA extracted from the labelled sediment (●) and the unlabelled controls (□). Experimental details are given in the Methods section. Numbers on x-axis indicate PLFAs as in Fig. 1.

important biogeochemical process that limits the release of methane, a greenhouse gas, from anaerobic environments⁵. Methane-oxidizing bacteria are extremely active in Lake Loosdrecht sediments, accounting for ~80% of the sediment oxygen consumption²³. Second, signature PLFA biomarkers can readily distinguish type I mesophilic methane-oxidizing bacteria, which predominantly contain a series of 16-carbon mono-unsaturated PLFAs, from type II, which contain 18-carbon mono-unsaturated PLFAs⁵. PLFA profiles are also useful for identifying specific genera within the type I subdivision²⁴. Third, methane-oxidizing bacteria are unique in their use of methane as a carbon and energy source⁵. Carbon provided as ¹³C-methane will therefore specifically label methane-oxidizing bacterial populations.

After labelling with ¹³C-methane, 16-carbon PLFAs rather than 18-carbon PLFAs showed a clear enrichment in carbon-isotopic signal (Fig. 3), suggesting that type I methanotrophs dominated methane oxidation at our freshwater site. Furthermore, the high amounts of label incorporated into 16:1 compared to 14:0 and 16:0 PLFAs suggests that the dominant organisms may belong to the type I genera *Methylobacter* and *Methylomicrobium*²⁴. Naturally occurring biogenic methane has a very depleted signal²⁵, and bacteria using this methane display its signal in their biomass^{26,27}. Even though we used isotopically rather heavy methane in our control incubations, the methanotrophic PLFAs that received most label were also depleted in the control incubations (Fig. 3c, squares), underlining the importance of type I methanotrophs in these sediments.

Our results clearly demonstrate the power of this new approach, which could potentially provide a wealth of information on many of the natural microbial populations involved in the main biogeochemical cycles. In principle, all processes that involve the biological consumption and incorporation of carbon by organisms could be studied using label additions, including autotrophic as well as heterotrophic processes. As shown in this study, ¹³C-labelling greatly extends the use of PLFA analysis in natural environments. As the full spectrum of the labelled PLFA can be used, researchers are no longer restricted to studying specific biomarkers that are only found in a limited number of genera. In addition, estimates of growth rates and yields of functional populations may be obtained in labelling studies, as polar lipid synthesis is closely linked to growth in bacteria²⁸. By directly linking activity with identity, this approach can greatly extend our knowledge about how the huge diversity of the microbial world relates to microbial processes in nature. □

Methods

¹³C-acetate labelling. In March 1995, uniformly labelled ¹³C-acetate (100 or 10 µM final concentration, 99% ¹³C) was injected at 0.5-cm intervals into small sediment cores sampled at the Tamar mud flat and triplicate cores were incubated for 0 and 8 h at *in situ* temperature (11 °C). All label was consumed after 8 h (determined as in ref. 29). In August 1995, the effect of 20 mM molybdate on incorporation of ¹³C-acetate (100 µM final concentration) was determined in a sediment slurry during a 2 h incubation (22 °C). Percentage label incorporated into a specific PLFA was calculated as: % incorporation = 100((F_{ix} - F₀)/[PLFA]_{ix})/[¹³C-acetate], with both concentrations in µgC per ml sediment and the fraction of ¹³C before (t₀) and after (t_x) incubation as: $F = \frac{^{13}\text{C}/(^{13}\text{C} + ^{12}\text{C})}{R/(R + 1)}$. The carbon isotope ratio (R) was derived from the measured δ¹³C values (in ‰) as: $R = (\delta^{13}\text{C}/1000 + 1)_{R_{\text{VPDB}}}$, with $R_{\text{VPDB}} = 0.0112$.

¹³C-methane labelling. Sediment cores were sampled at Lake Loosdrecht, The Netherlands (October 1997, *in situ* temperature 10 °C), and the top 4 cm of cores was transferred to a diffusion chamber²³ consisting of a lower compartment, through which anoxic, methane-saturated artificial pore water was pumped (containing 1.2 mM methane and 1.0 mM ammonium, reflecting *in situ* concentrations and resulting in natural fluxes of methane²³), and an upper compartment that contained the sediment. Compartments were separated by a 0.02-µm filter that allowed methane and other constituents to diffuse into the

sediment layer. All sediments were pre-incubated for 7 d with non-labelled methane (~30‰, determined by gas chromatography-isotope ratio monitoring mass spectroscopy). The isotopic ratio in the labelled incubations was subsequently raised to 1,350‰. All diffusion chambers (triplicates for both treatments) were incubated for another 7 d after which the top 5 mm of the sediment was analysed for PLFA. Incorporation of ¹³C into PLFAs was calculated as: incorporation = (F_{ix} - F₀)/[PLFA]_{ix} (definitions as above).

PLFA analysis. PLFA were extracted by a modified Bligh and Dyer method, fractionated on silica columns and trans-methylated to fatty acid methyl esters by mild alkaline methanolysis¹². Concentrations were determined by gas chromatography-flame ionization detection and δ¹³C values were measured on a Finnigan Delta-S isotope ratio monitoring mass spectrometer coupled to a gas chromatograph via a Finnigan Type II combustion interface^{27,30}. Identification of PLFA is based both on comparison of retention times with known reference compounds and on mass spectral analysis. All PLFA analyses were performed on two analytical columns differing in polarity.

Received 8 December 1997; accepted 10 February 1998.

- Amann, R. I., Ludwig, W. & Schleifer, K.-H. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**, 143–169 (1995).
- Teske, A., Waver, C., Muyzer, G. & Ramsing, N. B. Distribution of sulfate-reducing bacteria in a stratified fjord (Mariager Fjord, Denmark) as evaluated by most-probable-number counts and denaturing gradient gel electrophoresis of PCR-amplified ribosomal DNA fragments. *Appl. Environ. Microbiol.* **62**, 1405–1415 (1996).
- Jørgensen, B. B. Mineralization of organic matter in the sea bed. The role of sulphate reduction. *Nature* **296**, 643–645 (1982).
- Sørensen, J., Christensen, D. & Jørgensen, B. B. Volatile fatty acids and hydrogen as substrates for sulfate-reducing bacteria in anaerobic marine sediment. *Appl. Environ. Microbiol.* **42**, 5–11 (1981).
- Hanson, R. S. & Hanson, T. E. Methanotrophic bacteria. *Microbiol. Rev.* **60**, 439–471 (1996).
- Taylor, J. & Parkes, R. J. The cellular fatty acids of the sulphate-reducing bacteria, *Desulfobacter* sp., *Desulfobulbus* sp. and *Desulfovibrio desulfuricans*. *J. Gen. Microbiol.* **129**, 3303–3309 (1983).
- Dowling, N. J. E., Widdel, F. & White, D. C. Phospholipid ester-linked fatty acid biomarkers of acetate-oxidizing sulphate-reducers and other sulphide-forming bacteria. *J. Gen. Microbiol.* **132**, 1815–1825 (1986).
- Kohring, L. L. *et al.* Comparison of phylogenetic relationships based on phospholipid fatty acid profiles and ribosomal RNA sequence similarities among dissimilatory sulfate-reducing bacteria. *FEMS Microbiol. Lett.* **119**, 303–308 (1994).
- Parkes, R. J., Gibson, G. R., Mueller-Harvey, I., Buckingham, W. J. & Herbert, R. A. Determination of the substrates for sulphate-reducing bacteria within marine and estuarine sediments with different rates of sulphate reduction. *J. Gen. Microbiol.* **135**, 175–187 (1989).
- Wellsbury, P. & Parkes, R. J. Acetate bioavailability and turnover in an estuarine sediment. *FEMS Microbiol. Ecol.* **17**, 85–94 (1995).
- Oremland, R. S. & Capone, D. G. Use of "specific" inhibitors in biogeochemistry and microbial ecology. *Adv. Microb. Ecol.* **10**, 285–383 (1988).
- Guckert, J. B., Antworth, C. P., Nichols, P. D. & White, D. C. Phospholipid ester-linked fatty acid profiles as reproducible assays for changes in prokaryotic community structure of estuarine sediments. *FEMS Microbiol. Ecol.* **31**, 147–158 (1985).
- Ratledge, C. & Wilkinson, S. G. *Microbial lipids* (Academic, New York, 1988).
- Widdel, F. & Pfennig, N. A new anaerobic, sporing, acetate-oxidizing, sulfate-reducing bacterium, *Desulfotomaculum* (emend.) *acetoxidans*. *Arch. Microbiol.* **112**, 119–122 (1977).
- Widdel, F. & Pfennig, N. Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids. I. Isolation of new sulfate-reducing bacteria enriched with acetate from saline environments. Description of *Desulfobacter postgatei* gen. nov., sp. nov. *Arch. Microbiol.* **129**, 395–400 (1981).
- Keith, S. M., Herbert, R. A. & Harfoot, C. G. Isolation of new types of sulfate-reducing bacteria from estuarine and marine sediments using chemostate enrichments. *J. Appl. Bacteriol.* **53**, 29–33 (1982).
- Laanbroek, H. J. & Pfennig, N. Oxidation of short-chain fatty acids by sulfate-reducing bacteria in freshwater and in marine sediments. *Arch. Microbiol.* **128**, 330–335 (1981).
- Widdel, F. New types of acetate-oxidizing, sulfate-reducing *Desulfobacter* species, *D. hydrogenophilus* sp. nov., *D. latus* sp. nov., and *D. curvatus* sp. nov. *Arch. Microbiol.* **148**, 286–291 (1987).
- Devereux, R. & Mundfrom, G. W. A phylogenetic tree of 16S rRNA sequences from sulfate-reducing bacteria in a sandy marine sediment. *Appl. Environ. Microbiol.* **60**, 3437–3439 (1994).
- Waver, C. & Muyzer, G. Genetic diversity of *Desulfovibrio* spp. in environmental samples analyzed by denaturing gradient gel electrophoresis of [NiFe] hydrogenase gene fragments. *Appl. Environ. Microbiol.* **61**, 2203–2210 (1995).
- Lillebæk, R. Application of antisera raised against sulfate-reducing bacteria for indirect immunofluorescent detection of immunoreactive bacteria in sediment from the German Baltic Sea. *Appl. Environ. Microbiol.* **61**, 3436–3442 (1995).
- Devereux, R., Winfrey, M. R., Winfrey, J. & Stahl, D. A. Depth profile of sulfate-reducing bacterial ribosomal RNA and mercury methylation in an estuarine sediment. *FEMS Microbiol. Ecol.* **20**, 23–31 (1996).
- Sinke, A. J. C., Cottaar, F. H. M., Buis, K. & Keizer, P. Methane oxidation by methanotrophs and its effects on the phosphate flux over the sediment-water interface in an eutrophic lake. *Microb. Ecol.* **24**, 259–269 (1992).
- Bowman, J. P., Sly, L. I. & Stackelbrandt, E. The phylogenetic position of the family methylococcaceae. *Int. J. Sys. Bacteriol.* **45**, 182–185 (1993).
- Whiticar, M. J. & Faber, E. Methane oxidation in sediment and water column environments: Isotope evidence. *Org. Geochem.* **10**, 759–768 (1985).
- Summons, R. E., Jahnke, L. J. & Roksandic, Z. Carbon isotopic fractionation in lipids from methanotrophic bacteria: relevance for interpretation of the geochemical record of biomarkers. *Geochim. Cosmochim. Acta* **58**, 2853–2863 (1994).
- Freeman, K. H., Hayes, J. M., Trendel, J.-M. & Albrecht, P. Evidence from carbon isotope measurements for diverse origins of sedimentary hydrocarbons. *Nature* **343**, 254–256 (1990).
- Moriarty, D. J. W., White, D. C. & Wassenberg, T. J. A convenient method for measuring rates of phospholipid synthesis in seawater and sediments: Its relevance to the determination of bacterial productivity and the disturbance artefacts introduced by measurements. *J. Microbiol. Meth.* **3**, 321–330 (1985).
- de Graaf, W., Wellsbury, P., Parkes, R. J. & Cappenberg, T. E. Comparison of acetate turnover in

methanogenic and sulfate-reducing sediments by radiolabeling and stable isotope labeling and by use of specific inhibitors: Evidence for isotopic exchange. *Appl. Environ. Microbiol.* **62**, 772–777 (1996).
30. Pel, R. *et al.* Stable-isotope analysis of a combined nitrification-denitrification sustained by thermophilic methanotrophs under low-oxygen conditions. *Appl. Environ. Microbiol.* **63**, 474–481 (1997).

Acknowledgements. We thank H. J. Laanbroek, J. J. Middelburg and K. Perry for comments on earlier versions of the manuscript; L.-A. Meyer-Reil and M. Köster for their help at Rügen, Germany; and S. P. Barnes for the enrichment of acetate-using *Desulfotomaculum* from Tamar sediments. This work was supported by the European Union Environment Programme and the UK Natural Environmental Research Council.

Correspondence and requests for materials should be addressed to H.T.S.B. at CECE/NIE (e-mail: boschker@cemo.nioo.knaw.nl).

The viscosity of liquid iron at the physical conditions of the Earth's core

Gilles A. de Wijs*, Georg Kresse†‡, Lidunka Vočadlo‡, David Dobson‡, Dario Alfè*, Michael J. Gillan* & Geoffrey D. Price‡

* Physics Department, Keele University, Keele, Staffordshire ST5 5BG, UK

† Institute for Theoretical Physics, Technical University of Vienna,

Wiedner Hauptstrasse 8-10/136, A-1040 Vienna, Austria

‡ Research School of Geological and Geophysical Sciences, Birkbeck College and University College London, Gower Street, London WC1E 6BT, UK

It is thought that the Earth's outer core consists mainly of liquid iron and that the convection of this metallic liquid gives rise to the Earth's magnetic field. A full understanding of this convection is hampered, however, by uncertainty regarding the viscosity of the outer core. Viscosity estimates from various sources span no less than 12 orders of magnitude^{1,2}, and it seems unlikely that this uncertainty will be substantially reduced by experimental measurements in the near future. Here we present dynamical first-principles simulations of liquid iron which indicate that the viscosity of iron at core temperatures and pressures is at the low end of the range of previous estimates—roughly 10 times that of typical liquid metals at ambient pressure. This estimate supports the approximation commonly made in magnetohydrodynamic models that the outer core is an inviscid fluid^{3–5} undergoing small-scale circulation and turbulent convection⁶, rather than large-scale global circulation.

In first-principles calculations, the system of interest is represented as a collection of atomic nuclei and electrons, and the total energy and the forces on all the nuclei are obtained by solving Schrödinger's equation to determine the electronic ground state. The calculations are based on density functional theory⁷, with the electronic exchange-correlation energy treated using either the local-density approximation or the more sophisticated generalized-gradient approximation. First-principles calculations on a large variety of materials, including transition metals, have been shown to give accurate predictions of both static and dynamic quantities (see, for example, ref. 8). There have already been extensive first-principles calculations on crystalline iron over the pressure range from zero up to Earth's-core values^{9,10}, and it has been shown that the generalized-gradient approximation reproduces the density of the hexagonal close-packed (h.c.p.) structure (the likely stable structure at high pressure¹¹) over the whole high-pressure range within ~6%; at ambient pressure the volume is overestimated slightly more (by ~9%). The equilibrium density and magnetic properties of the body-centred cubic (b.c.c.) structure of iron at ambient pressure are also accurately described.

To study a liquid with these methods, we must perform first-principles molecular dynamics. Car and Parrinello first showed how to do this by integrating Newton's equation of motion for the nuclei, with the forces calculated from the first-principles ground state¹². The first-principles molecular dynamics technique that they

pioneered uses the pseudopotential approach, in which only the valence electrons are explicitly represented, with the orbitals expanded in a plane-wave basis. Techniques of this kind have been widely used for studying liquid metals (see, for example, refs 8, 13, 14), and their accuracy in predicting both structural and dynamic properties is well established.

Our first-principles molecular dynamics simulations are based on ultrasoft pseudopotentials¹⁵ of the Vanderbilt type¹⁶, which allow a significant reduction in the computational effort with no loss of accuracy (the accuracy of the pseudopotentials will be demonstrated below). Electronic exchange-correlation is treated using the generalized-gradient approximation of Perdew *et al.*¹⁷. In achieving good accuracy for high-pressure iron, the choice of orbitals to be treated as valence orbitals is crucial. At ambient pressure, reasonable accuracy can be obtained by treating all atomic states up to and including 3p as core states, but this is not satisfactory here because the 3p states respond significantly at high pressures. By contrast, inclusion of 3s states in the atomic core gives insignificant errors, provided that nonlinear core corrections¹⁸ are included.

We have probed the accuracy of our pseudopotential calculations using a range of tests on solid iron, with 3p states and above included in the valence set. Our calculated values of the atomic volume and bulk modulus of the b.c.c. and h.c.p. phases of iron, and the magnetic moment of the b.c.c. phase at ambient pressure, are compared in Table 1 with experimental values^{19,20} and all-electron calculations⁹. The close agreement with the all-electron results indicates that the errors incurred by the pseudopotential approximation are very small, and the agreement with experiment confirms the accuracy of the generalized-gradient approximation. We tested the reliability of our calculations at high pressures by comparing our calculated pressure as a function of volume for the h.c.p. phase against experimental measurements¹¹ and previous all-electron predictions^{9,10}. We find that the discrepancy between our predicted pressures and the measured values is <3% over the range 50–300 GPa, which covers most of the pressure range of the outer core. We have also used calculations of enthalpy as a function of pressure for the b.c.c. and h.c.p. phases of iron to predict the transition pressure between the two structures; our calculated value of ~10 GPa agrees with experimental values²¹ in the range 10–15 GPa and with the value of ~11 GPa given by all-electron linearized augmented planewave⁹ and linearized muffin-tin orbital¹⁰ calculations. Finally, we have compared our calculated zone-boundary phonon frequencies for the highly compressed f.c.c. phase with the all-electron results of Söderlind *et al.*¹⁰. The discrepancies of <3% for all frequencies confirm again the accuracy of our pseudopotential methods.

Our first-principles molecular dynamics simulations of molten iron use essentially the same approximations as our calculations on the solid phases, and we expect to achieve similar accuracy. (For present purposes, the 3p states are not included explicitly, but their effect is correctly reproduced by a pairwise additive potential.) We performed simulations at thermodynamic states corresponding roughly to the core–mantle boundary (CMB) and the boundary between the molten outer core and the solid inner core (ICB). The temperature *T* in the Earth's core is uncertain: estimates of *T* at the ICB range from 4,000 to 8,000 K and at the CMB from 3,000 to

Table 1 Calculated and experimental properties of iron

Phase property	All electron (ref. 9)	Pseudopotential (this work)	Expt (refs 19, 20)
b.c.c. V_0 (Å ³)	11.4	11.55	11.80
b.c.c. K (GPa)	189	176	162–176
b.c.c. μ (μ_B per atom)	2.17	2.25	2.12
h.c.p. V_0 (Å ³)	10.2	10.4	11.2
h.c.p. K (GPa)	291	290	208

Comparison of present pseudopotential calculations with all-electron calculations and experimental values for atomic volume V_0 , bulk modulus K and magnetic moment μ of b.c.c. and h.c.p. iron.

4,500 K (ref. 22). The pressure in the core is accurately known (330 GPa at the ICB)²². Initially we assumed that the core lies on an adiabat whose temperature at the ICB pressure is 6,000 K. Using the data compiled by Anderson and Ahrens²³ for the equation of state of liquid iron, we then find that the density ρ at the ICB state is $1.33 \times 10^4 \text{ kg m}^{-3}$, and that T and ρ at the CMB state are 4,300 K and $1.07 \times 10^4 \text{ kg m}^{-3}$, respectively. In order to allow for the uncertainty in T , simulations at these two states were supplemented by a further simulation at the CMB value of ρ but at the lower temperature of 3,500 K. The simulations used a periodically repeated 64-atom system (for possible system-size effects, see below), and Brillouin-zone sampling was performed at the Γ -point only. All three simulation runs had a duration of ~ 2 ps after equilibration. They showed good stability, with total energy drift corresponding to temperature changes of only a few tens of degrees per picosecond.

An important check on the realism of our simulations can be made by comparing the calculated pressure of the liquid with the Anderson and Ahrens data²³. As shown by the results in Table 2, the calculated and experimental pressures differ by $<10\%$ for the ICB state and even less for the CMB state. Our simulations show that the structure of high-pressure molten iron, as described by the radial distribution function $g(r)$, is close packed, with a coordination number of slightly over 12. (Detailed $g(r)$ results will be reported elsewhere²⁴.) This means that the liquid structure closely resembles that of the h.c.p. crystal.

We obtain an estimate for the shear viscosity η of molten iron from our simulations by using its connection with the self-diffusion coefficient D . (The direct calculation of η from first-principles simulations is theoretically possible, but has not yet been achieved for any liquid, because of the very long simulations needed to obtain useful statistical accuracy.) We calculate D via the Einstein relation:

$$\langle \Delta r(t)^2 \rangle \rightarrow B + 6Dt \quad (1)$$

where B is a constant. Equation (1) tells us that the mean square distance $\langle \Delta r(t)^2 \rangle$ travelled by any atom in time t becomes linear in t at long times. In all our simulations, we find that $\langle \Delta r(t)^2 \rangle$ settles down to a linear dependence on t in <0.5 ps, so that values for D are straightforward to extract. Our values of D at the CMB and ICB states (Table 2) are of the same order as the diffusion coefficients of typical liquid metals at ambient pressure²⁵.

We now obtain the viscosity via the Stokes–Einstein equation:

$$D\eta = k_B T / 2\pi a \quad (2)$$

This equation from the theory of brownian motion gives an exact relation between the diffusion coefficient D of a brownian particle of diameter a and the viscosity η of the surrounding liquid (k_B is the Boltzmann constant). It is often applied also to the diffusion of atoms in liquids, in which case the relation is only approximate. However, its empirical accuracy when applied to the diffusion of atoms is very good. If a is taken to be the nearest-neighbour distance in the solid, then predictions of η using measured values of D agree with experimental data to within 40% for a wide range of liquid metals. Simulation studies on the hard-sphere liquid²⁶ demonstrate that the accuracy of the Stokes–Einstein relation remains very good even at high pressures where the free volume becomes small. Taking

a from our calculated pressure–volume relation for the h.c.p. solid, we obtain the η values for molten iron under Earth’s-core conditions reported in Table 2. Our calculated viscosities are roughly ten times those of typical liquid metals at ambient pressure.

In assessing the reliability of our calculations, we note first that our simulations were necessarily done on a rather small system. System size errors in calculated transport coefficients have been exhaustively studied for simple systems such as the Lennard–Jones and hard-sphere liquids. (These are highly relevant, because the present simulations demonstrate that high-pressure liquid iron closely resembles these liquids.) In such systems, the calculated D depends only weakly on the system size, and the simulation data of ref. 27 indicate that D calculated with 64 atoms will underestimate the true value by $\sim 20\%$. This means that our η values obtained from the Stokes–Einstein relation are probably too high by $\sim 20\%$. We have already noted that the Stokes–Einstein relation itself should be good to 40% or better, so the technical error in our η values should be no worse than 60%. Turning now to uncertainties about the temperature T in the core, we note from Table 2 that the CMB value of η increases by $\sim 25\%$ as T goes from 4,300 to 3,500 K. Assuming an Arrhenius temperature dependence, this means that if T at the CMB were as low as 3,000 K, η would be 1.8×10^{-2} Pa s. Taking all these uncertainties together, we believe that the overall error is unlikely to be worse than a factor of three.

We also note that the Earth’s core does not consist of pure iron, but contains light impurities (carbon, oxygen, silicon or sulphur) at a total concentration of up to 10%. The effect of these impurities on the viscosity under Earth’s-core conditions is not known, but they seem very unlikely to affect it substantially, given their low concentration. (The influence of dissolved sulphur on the viscosity of liquid iron at much higher concentrations has been studied in ref. 28.)

Our overall conclusion is that the shear viscosity of liquid iron in the Earth’s core is $\sim 1.5 \times 10^{-2}$ Pa s, with an uncertainty of a factor of three. This is at the low end of the range of previous estimates, though it is similar to values suggested by some earlier work (see, for example, ref. 1). It is very much lower than apparent viscosity values deduced from geodetic and seismic measurements, but it is recognized that because of their timescale the latter may be influenced by dissipative processes other than shear viscosity^{2,29}. A low viscosity has important implications for the geodynamic description of the outer core. The nonlinear equations of magnetohydrodynamics cannot be solved exactly, and simplified linear approximations are often used. Depending on these approximations, contributions to the forces acting on the fluid in the outer core, such as Coriolis, Lorentz and viscous, may be treated as small or negligible. The relative importance of the viscous and Coriolis forces is characterized by the Ekman number given by $\text{Ek} = \eta / (\rho \Omega L^2)$, where ρ is the fluid density, Ω is the Earth’s rotational velocity, and L is the fluid thickness. With $\eta = 0.015$ Pa s, $\rho = 1.06 \times 10^4 \text{ kg m}^{-3}$, $\Omega = 7.3 \times 10^5 \text{ rad s}^{-1}$ and $L = 2,000$ km, this gives $\text{Ek} = 4 \times 10^{-15}$, indicating that viscous forces are indeed negligible when compared with the Coriolis force. This supports models based on the assumption that the viscosity of the core is negligible^{3–5}, and favours a picture in which the core is in a state of small-circulation turbulent convection, in contrast to models having a viscosity of $\sim 10^7$ Pa s which imply a coherent pattern on a much larger scale, comparable with the core radius^{6,30}. □

Table 2 Calculated and experimental properties of iron at Earth’s-core conditions

	T (K)	ρ (10^4 kg m^{-3})	ρ_{exp} (GPa)	ρ_{calc} (GPa)	D ($10^{-8} \text{ m}^2 \text{ s}^{-1}$)	η (10^{-2} Pa s)
ICB	6,000	1.33	330	358	0.5	1.3
CMB	4,300	1.07	135	132	0.4	1.2
CMB	3,500	1.07		125	0.3	1.5

Shown are the calculated and experimental pressures ρ_{calc} and ρ_{exp} , and calculated diffusion coefficient D and viscosity η of molten iron at thermodynamic states corresponding to the inner-core boundary (ICB) and the core–mantle boundary (CMB). Values of temperature T and density ρ are explained in the text.

Received 29 January; accepted 26 February 1998.

- Poirier, J. P. Transport properties of liquid metals and viscosity of the Earth’s core. *Geophys. J.* **92**, 99–105 (1988).
- Secco, R. A. in *Mineral Physics and Crystallography: A Handbook of Physical Constants* (ed. Ahrens, T. J.) 218 (American Geophysical Union, 1995).
- Glatzmaier, G. A. & Roberts, P. H. A three-dimensional self-consistent computer simulation of a geomagnetic field reversal. *Nature* **377**, 203–209 (1995).
- Buffett, B. A., Huppert, H. E., Lister, J. R. & Woods, A. W. On the thermal evolution of the Earth’s core. *J. Geophys. Res.* **101**, 7989–8006 (1996).
- Merrill, R. T. & McFadden, P. L. Dynamo theory and paleomagnetism. *J. Geophys. Res.* **100**, 317–326 (1995).

6. Melchior, P. *The Physics of the Earth's Core* (Pergamon, Oxford, 1986).
7. Jones, R. O. & Gunnarsson, O. The density functional formalism, its applications and prospects. *Rev. Mod. Phys.* **61**, 689–746 (1989).
8. Kresse, G. & Hafner, J. *Ab initio* molecular dynamics for open-shell transition metals. *Phys. Rev. B* **48**, 13115–13118 (1993).
9. Stixrude, L., Cohen, R. E. & Singh, D. J. Iron at high pressure: Linearized-augmented-plane-wave computations in the generalized-gradient approximation. *Phys. Rev. B* **50**, 6442–6445 (1994).
10. Söderlind, P., Moriarty, J. A. & Wills, J. M. First-principles theory of iron up to earth-core pressures: Structural, vibrational, and elastic properties. *Phys. Rev. B* **53**, 14063–14072 (1996).
11. Mao, H. K., Wu, Y., Chu, L. C., Shu, J. F. & Jephcoat, A. P. Compression of iron to 300 GPa and Fe_{0.8}Ni_{0.2} alloy to 260 GPa: Implications for composition of the core. *J. Geophys. Res.* **95**, 21737–21742 (1990).
12. Car, R. & Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **55**, 2471–2474 (1985).
13. Holender, J. M., Gillan, M. J., Payne, M. C. & Simpson, A. D. The static, dynamic and electronic properties of liquid gallium studied by first-principles simulation. *Phys. Rev. B* **52**, 967–975 (1995).
14. Kresse, G. & Hafner, J. *Ab initio* molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium. *Phys. Rev. B* **49**, 14251–14269 (1994).
15. Moroni, E., Kresse, G., Furthmüller, J. & Hafner, J. Pseudopotentials applied to magnetic Fe, Co and Ni: From atoms to solids. *Phys. Rev. B* (submitted).
16. Vanderbilt, D. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B* **41**, 7892–7895 (1990).
17. Perdew, J. et al. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B* **46**, 6671–6687 (1992).
18. Louie, S. G., Froyen, S. & Cohen, M. L. Nonlinear ionic pseudopotentials in spin-density-functional calculations. *Phys. Rev. B* **26**, 1738–1742 (1982).
19. Ahrens, T. J. (ed.) *Mineral Physics and Crystallography: A Handbook of Physical Constants* (American Geophysical Union, 1995).
20. Lonzarich, G. G. in *Electrons at the Fermi Surface* (ed. Springford, M.) 257 (Cambridge Univ. Press, 1980).
21. Anderson, O. L. The phase diagram of iron and the temperature of the inner core. *J. Geomag. Geoelectr.* **45**, 1235–1248 (1993).
22. Poirier, J. P. *Introduction to the Physics of the Earth's Interior* (Cambridge Univ. Press, 1991).
23. Anderson, W. W. & Ahrens, T. J. An equation of state for liquid iron and implications for the Earth's core. *J. Geophys. Res.* **99**, 4273–4284 (1994).
24. Vočadlo, L., de Wijs, G. A., Kresse, G., Gillan, M. J. & Price, G. D. First principles calculations of crystalline and liquid iron at Earth's core conditions in *Solid-State Chemistry—New Opportunities from Computer Simulations* (eds Mackrodt, W. & Catlow, R.) (Faraday Discussion No. 106, Royal Society of Chemistry, London, 1997).
25. Shimoi, M. & Itami, T. *Atomic Transport in Liquid Metals* (Trans Tech, Aedermannsdorf, 1986).
26. Alder, B. J., Gass, D. M. & Wainwright, T. E. The transport coefficients for a hard-sphere fluid. *J. Chem. Phys.* **53**, 3813–3826 (1970).
27. Erpenbeck, J. J. & Wood, W. W. Molecular dynamics calculation of the velocity autocorrelation function: hard-sphere results. *Phys. Rev. A* **32**, 412–422 (1985).
28. LeBlanc, G. E. & Secora, R. A. Viscosity of an Fe-S liquid up to 1300 °C and 5 GPa. *Geophys. Res. Lett.* **23**, 213–216 (1996).
29. Anderson, D. L. Bulk attenuation in the Earth and viscosity of the core. *Nature* **285**, 204–207 (1980).
30. Officer, C. B. A conceptual model of core dynamics and the Earth's magnetic field. *J. Geophys.* **59**, 89–97 (1986).

Acknowledgements. The simulations were run on the Cray T3D at Edinburgh Parallel Computer Centre with support from the High Performance Computing Initiative. We thank E. Moroni, J. Hafner and J. Brodholt for discussions.

Correspondence and requests for materials should be addressed to M.J.G. (e-mail: pha71@keele.ac.uk).

Reproductive cessation in female mammals

Craig Packer*, Marc Tatar† & Anthony Collins‡

* Department of Ecology, Evolution and Behavior, University of Minnesota, 1987 Upper Buford Circle, St Paul, Minnesota 55108, USA

† Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912, USA

‡ Gombe Stream Research Centre, Box 185, Kigoma, Tanzania

In female mammals, fertility declines abruptly at an advanced age. The human menopause is one example, but reproductive cessation has also been documented in non-human primates, rodents, whales, dogs, rabbits, elephants and domestic livestock^{1–3}. The human menopause has been considered an evolutionary adaptation^{4–7}, assuming that elderly women avoid the increasing complications of continued childbirth to better nurture their current children and grandchildren. But an abrupt reproductive decline might be only a non-adaptive by-product of life-history patterns. Because so many individuals die from starvation, disease and predation, detrimental genetic traits can persist (or even be favoured) as long as their deleterious effects are delayed until an advanced age is reached, and, for a given pattern of mortality, there should be an age by which selection would be too weak to

prevent the onset of reproductive senescence^{4,5,8}. We provide a systematic test of these alternatives using field data from two species in which grandmothers frequently engage in kin-directed behaviour. Both species show abrupt age-specific changes in reproductive performance that are characteristic of menopause. But elderly females do not suffer increased mortality costs of reproduction, nor do post-reproductive females enhance the fitness of grandchildren or older children. Instead, reproductive cessation appears to result from senescence.

Age-specific rates of mortality and maternity are broadly similar in olive baboons and African lions (Fig. 1). In baboons, female mortality reaches a minimum between 4 and 5 years of age and then accelerates; no female has survived beyond 27 years of age (Fig. 1a). But the maternity rate remains constant until 21 years of age, when it decreases (Fig. 1c). In female lions, mortality reaches a minimum between 3 and 4 years and then rapidly accelerates; no lioness has survived beyond 17 years of age (Fig. 1b). Lion maternity rates remain relatively constant before decreasing at 14 years of age (Fig. 1d).

Infant survival declines rapidly when female baboons reach 21 years (Fig. 2a), which is the age at which pregnancies are more likely to end in miscarriage (Fig. 2b). Baboons also show striking changes in menstrual cycles with advancing age. The length of the cycle remains constant until about 23 years of age (Fig. 3a), when cycles also become more irregular (Fig. 3b). Fertility decreases at 23 years of age and essentially ceases at 24 (Fig. 3c). Because of low fertility, elderly females experience a peak number of cycles during their 24th year (comparable to the period of 'adolescent sterility'), but cycling diminishes thereafter until ceasing altogether (Fig. 3d). Constant cycling is known to lower the age of menopause in women⁹, and one female baboon stopped cycling at 20 years and remained acyclic until her death aged 26. She had shown unusually low fertility throughout life and experienced 127 menstrual cycles by her 20th birthday, whereas all other 20-year-old females averaged only 50 cycles (range 35–98, $n = 23$).

Less detailed data are available on the precise reproductive performance of ageing female lions (see Methods). The survival of lion cubs remains constant with maternal age (Fig. 2c), but litter size declines markedly at 14 years (Fig. 2d). Falling litter size cannot account for the overall decline in maternity rates between prime- and old-aged females (Fig. 1d), so elderly lions must also experience a reduction in cycling and/or fertility.

Are these declines adaptive? Baboons and lions are good candidates for an 'adaptive menopause' because both species engage in a high degree of kin-directed cooperative behaviour. Typical of cercopithecine monkeys, baboon mothers help to determine their daughters' dominance rank, and they also groom and intervene on behalf of their descendent kin¹⁰. Female lions participate in joint territorial defence^{11–13} and, when raising cubs communally, nurse their daughters' cubs as often as their own¹⁴.

The adaptive menopause hypotheses assume that post-reproductive females actively enhance the fitness of their prior offspring^{4–6}. This benefit would be lost by her death in the 'risky childbirth' hypothesis⁴ or by the intensive demands of infant care in the 'opportunity costs' hypothesis⁷. However, neither the survival of grandchildren nor the reproductive performance of adult daughters is improved in the predicted manner (Fig. 4). Lion cubs only show higher survival when their grandmothers are reproductively active: elderly female lions only engage in allomothering while tending their own cubs¹⁴. The typical interbirth interval is two years in both species and, whereas infant survival is highly dependent on maternal survival (Fig. 5), juvenile survival is unaffected by either the mother's survival or her subsequent reproduction (Fig. 5). Thus mothers do not invest in subsequent offspring until their prior brood has been successfully 'fledged' and continued reproduction inflicts no opportunity costs.

In both species, breeding females showed similar survival to non-

breeders (Fig. 6a, b), and a proportional hazards analysis using parity as the time-dependent covariate found mortality among reproductive females to be only 84% of that of non-reproductives in baboons ($c_2 = 24.06$, $\exp(b) = 0.843$) and 72% in lions ($c_2 = 70.65$, $\exp(b) = 0.719$). Without controlled experiments, we cannot distinguish these results from sample heterogeneity, as good

health may promote survival as well as reproduction (see ref. 15). Despite the possibility of heterogeneity, our data can still show whether reproductive costs increase with age. However, the relative risk of reproduction in both species remains constant with age (although it is possible that sample heterogeneity also changes with age¹⁶). Further, baboons showed no significant change in mortality

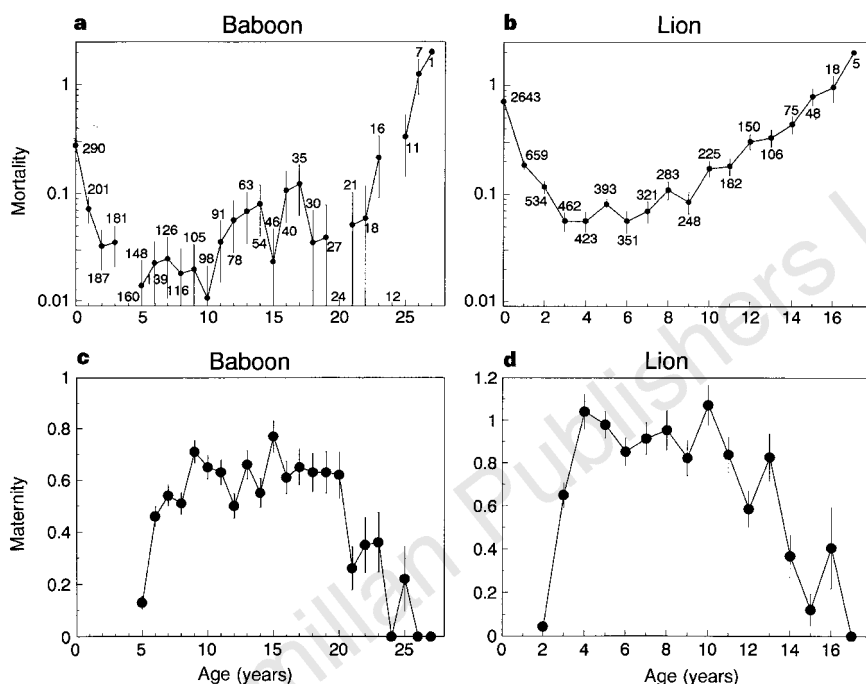


Figure 1 Age-specific mortality and maternity in female baboons and lions. Bars represent standard errors; numbers are sample size. **a, b.** Annual mortality is estimated from weekly intervals for baboons (**a**; 201 yearlings, 60.7% censored) and from monthly intervals for lions (**b**; 652 yearlings, 18.4% censored). Mortality rate = $d_i / b_i(n_i - 0.5d_i)$, where d_i is the number dying in the i th interval, b_i is the width of the i th interval, and n_i is the number at risk during the i th interval²⁶. No

deaths were observed in three baboon age classes, and $\ln(\text{mortality rate})$ is undefined. Infant mortality in lions includes all known females as well as unsexed cubs whereas later ages only include known females. Gross maternity for baboons (**c**) and lions (**d**) is the number of live offspring produced at each age and is estimated by the total number of live offspring born to age class i divided by the midpoint number of females in age class i .

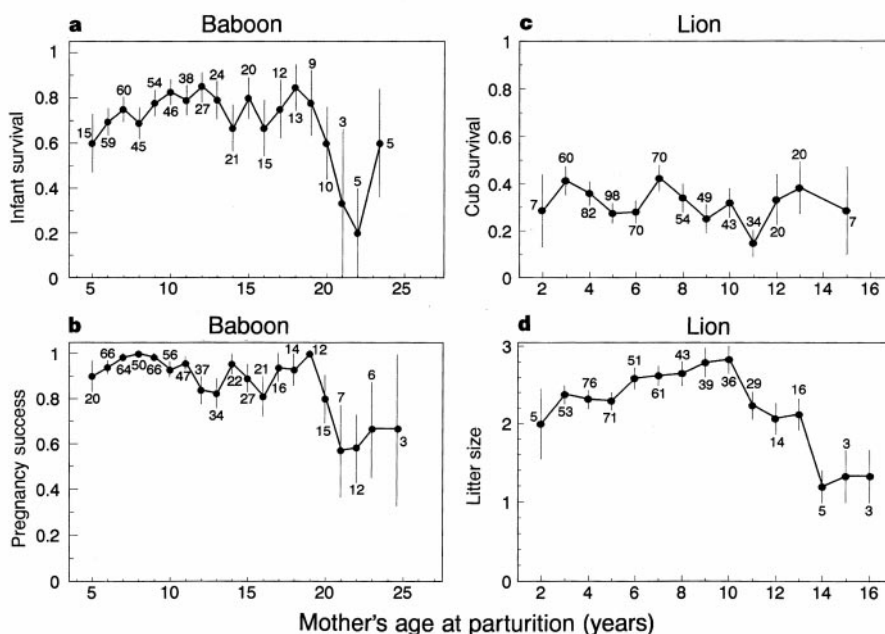


Figure 2 Age-specific aspects of female reproduction. **a.** First-year infant survival in baboons. **b.** Proportion of baboon pregnancies that ended in live birth. **c.** Average first-year survival for lion litters. **d.** Average lion litter size. Bars represent standard errors, and numbers indicate number of infants, pregnancies or litters.

All measures except lion-cub survival show significant heterogeneity with maternal age (ANOVA, $P < 0.01$) and a significant ($P < 0.01$) decline at advanced ages. Because maternal survival influences infant survival (see Fig. 5), data in **a** and **c** are restricted to mothers who survived at least one year post partum.

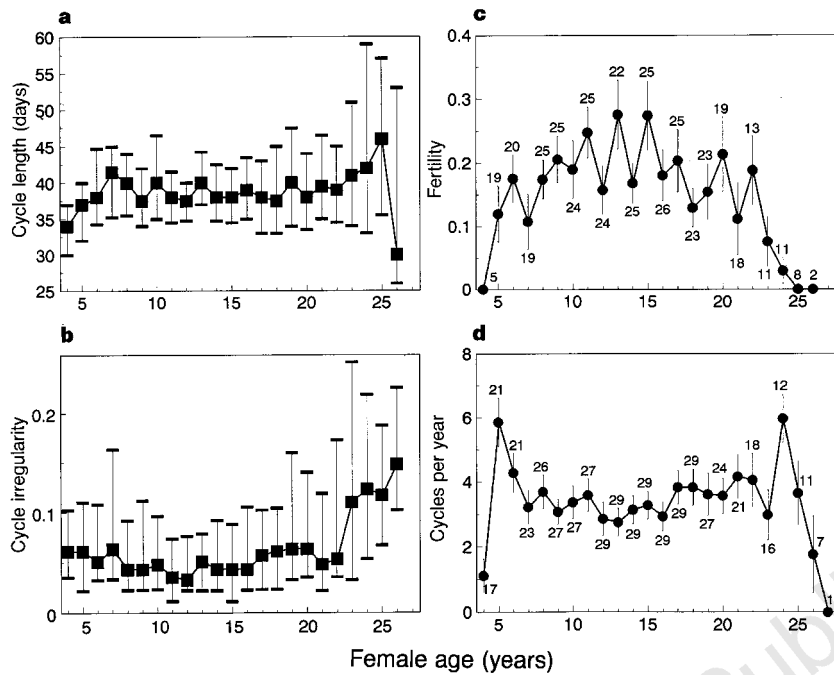


Figure 3 Age-specific aspects of baboon menstrual cycles. **a**, Median (squares) and quartile (bars) cycle lengths at each age. **b**, Median and quartile 'irregularity' of menstrual cycles (see text). **c**, Mean proportion of cycles resulting in pregnancy (plotted with standard errors and the number of cycling females at each age). **d**, Annual cycling rates (with standard errors and the total number of females of each age). All four measures show significant heterogeneity with age (Kruskal-Wallis or ANOVA, $P < 0.01$). Cycles become significantly more irregular while fertility declines at advanced ages (both with $P < 0.001$).

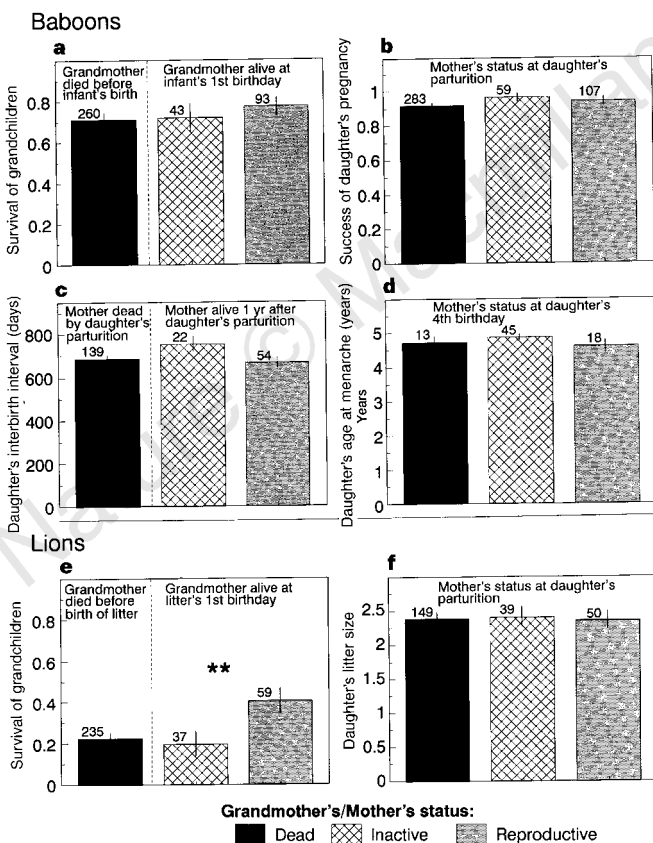


Figure 4 Effects of female survival and reproduction on descendants' productivity. Females were 'inactive' if they had not reproduced during the previous 12 months and for 6 months thereafter; otherwise they were 'reproductive'. **a-d**, Baboons. **a**, Proportion of infant grandchildren that survived their first year. **b**, Proportion of daughters' pregnancies that reached full term. **c**, Interval from birth of surviving grandchild until the next live birth. **d**, Age at which daughters experienced their first sexual swelling. **e, f**, Lions. **e**, Proportion of grandchildren in each litter that survived their first year. **f**, Daughters' average litter size. The only significant effect was in lions, where grandchildren of reproductively active grandmothers enjoyed higher survival (ANOVA, $F = 6.14$, d.f. = 2,328, $P = 0.0026$). Double asterisks indicate $P < 0.01$.

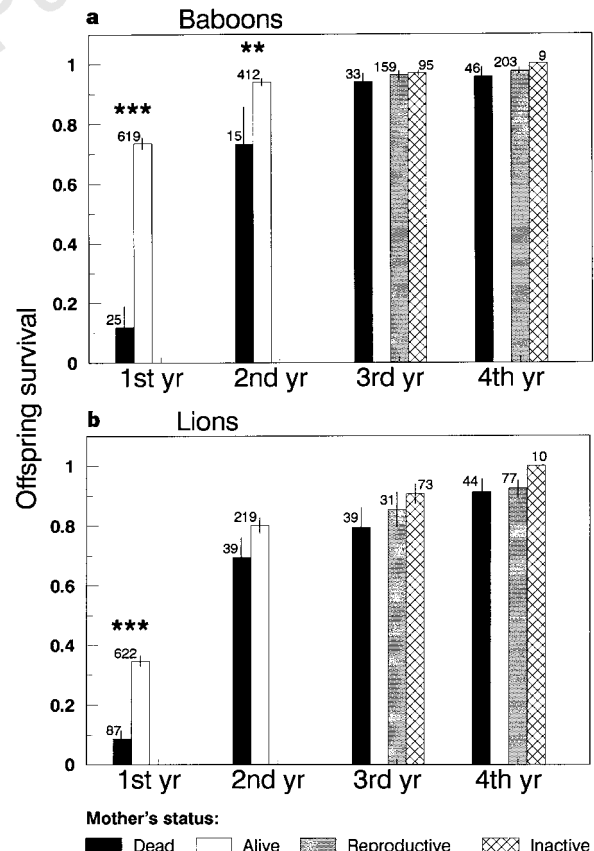


Figure 5 Effects of female survival and subsequent reproduction on offspring survival. **a**, Baboons and **b**, lions. Survival was significantly lower for animals orphaned during either their first year (baboons: $\chi^2 = 17.79$, $P = 0.0000$; lions: $z = 4.467$, $P = 0.0000$) or their second year (baboons: $\chi^2 = 6.72$, $P = 0.0095$). The typical interbirth interval for both species is two years, so effects of subsequent reproduction could be measured only in older juveniles. Mothers who gave birth before their prior offspring's second birthday were 'reproductive' for the juvenile's third year, otherwise they were 'inactive'. Juvenile survival was not influenced by the mothers' reproduction. Because of male dispersal in lions, 'juveniles' include only females. Double asterisks indicate $P < 0.001$; triple asterisks indicate $P < 0.0001$.

during pregnancy, parturition and the first year post partum (Fig. 6c), nor is there any evidence that such mortality increases with age. However, even if reproductive costs are higher than our measurements imply, both species fledge each brood before breeding again, so the effects of maternal mortality would be restricted to her youngest brood.

The rapid decline in reproductive performance of elderly females may result from low selection pressure on later ages⁵. If the force of natural selection is sufficiently weak at 21 and 14 years in baboons and lions, respectively, the observed reproductive declines could evolve through alleles with late-age-limited effects¹⁷. In this case, the difference in fitness between the observed cohort and a hypothetical non-menopausal cohort should be very small. We estimated this difference by calculating the population growth λ (the dominant eigen-value of the population projection matrix¹⁸) for each species using first the observed vital rates, and second the rates of a hypothetical cohort where fertility at advanced ages was the same as for younger females. Among baboons, the observed λ was 1.1329 compared with 1.1355 for a non-menopausal population. Among lions, the observed λ was 1.1970 compared with 1.1985 for the hypothetical population. Thus maintaining reproduction at advanced ages would make only a negligible contribution to fitness in either species. Natural selection therefore has little opportunity to oppose the evolution of reproductive senescence resulting from either pleiotropic alleles that confer benefits in early life or the accumulation of deleterious mutations with late-age-limited effects.

The age-specific reproductive performance of female mammals shows a pattern of constancy followed by sudden decline. Physiologically, declining fertility is caused by a deteriorating ovarian environment, and menopause results from the breakdown of neural pacemakers which depletes the finite store of oocytes^{1,19}. Senescence is predicted (and mortality senescence is generally observed; Fig. 1a, b) to begin at the age of first reproduction, and reproductive performance declines monotonically in many reptiles, birds and

invertebrates². But gradual reproductive senescence in mammals may be masked by extensive maternal care and small litter size. Young females may be physiologically capable of producing larger litters than they could successfully rear, but this capacity eventually declines until they can no longer produce typical numbers of offspring (Fig. 2b, d) before ceasing altogether (Fig. 3).

The timing of reproductive senescence will be set by both the female's age-specific life expectancy and the duration of infant dependency. Even in species in which reproduction inflicts no mortality, females will eventually reach an age where they would either die (Fig. 1a, b) or lose their vigour (Fig. 2a) before they could fledge additional offspring. At this point, any degeneration of reproductive performance will be unopposed by selection because no further fitness can be gained from continued parturition. However, selection will maintain somatic function long enough to rear any offspring that are dependent on the mother when she reaches menopause. Consequently, the time lag between reproductive and somatic senescence should be greatest in species in which infant dependency is most prolonged. Infant baboons suffer significant mortality when orphaned at 2 years of age (Fig. 5), and the females' life expectancy at 21 years is an additional 5.0 years; lion cubs are only vulnerable during their first year, and life expectancy at 14 years is only 1.8 years. Assuming a childhood dependency of 10 years and extrapolating from the observed relationship between mortality and maternity in baboons and lions, the expected lifespan of women who have reached 40 years (when reproduction starts to decline^{9,19,20}) would be 58–65 years. There is no consensus on mortality in pre-technological societies²¹, but prolonged infant dependency seems sufficient to account for a midlife menopause in human females (although increased reproductive costs would lead to an even younger menopause in any species in which interbirth intervals are shorter than the period of infant dependency). Infant dependency also explains why reproductive cessation is more striking in females than males: most mammalian infants

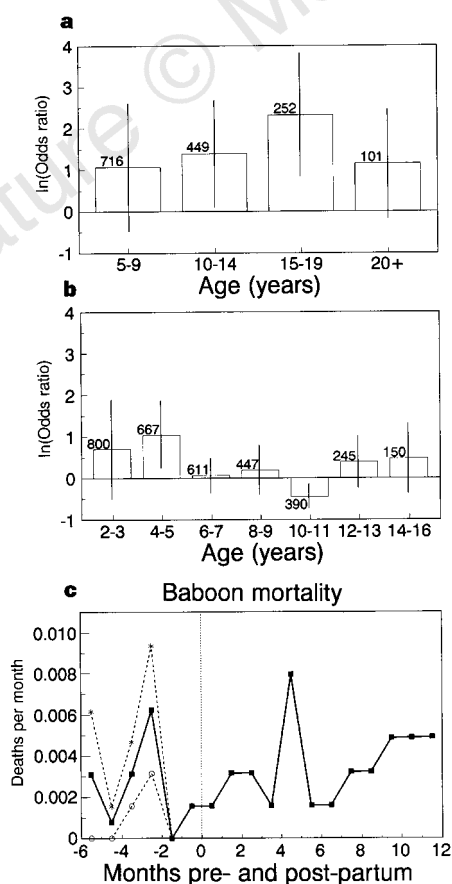


Figure 6 Costs of active reproduction. Logarithm of odds ratio (95% confidence interval) for survival at each age for reproductive females compared to non-reproductive females in baboons (**a**) and lions (**b**). Values are summed into multi-year classes to ensure non-zero frequencies; numbers refer to female years. Baboons aged 15–19 years were the only age class to show a significant deviation from zero ($\chi^2 = 7.08$, $P < 0.01$). **c**, Monthly mortality rates for female baboons throughout pregnancy and lactation. Post-partum mortality is based on the 635 live births of 160 females. 'Known pregnancy' (open circles) includes four females who were clearly pregnant at their death; 'possible pregnancy' (asterisks) includes these four plus 11 additional females who died within 6 months of their final cycle; filled squares indicate the midpoint between alternatives. Mortality did not vary across months in these data, nor during 12 months following miscarriage.

depend on the post-partum vigour and survival of the mother rather than of the father. □

Methods

Study populations. Olive baboons at Gombe National Park, Tanzania, have been studied since 1967 (ref. 22). These animals have been censused almost daily since 1972, and the demographic data include all birth dates, miscarriages, mortality and reproductive cycle state. Menarche typically occurs during the fourth year, so our analysis includes females that were born as early as 1963. Lions in Serengeti National Park and Ngorongoro Crater, Tanzania, have been monitored since 1966 and 1962, respectively²³. These animals are censused 2–10 times each month, and relevant demographic data include litter size, birth dates and mortality. Birth dates can be estimated from the body size of animals less than 2 years old, so we include females born as early as 1964 or 1960 in Serengeti and Ngorongoro, respectively. The reproductive state of the lions can be inferred only on the basis of observed births (and not all births are detected), so the lion analysis is restricted to cub productivity and female mortality.

Baboon menstrual cycles. Ovulatory cycle state is assessed from the female's perineum, which undergoes conspicuous phases of tumescence and detumescence and turns scarlet after conception²⁴. Analysis of 1,461 menstrual cycles is based on the 29 females of known age who lived for 18 years or more. Cycle length was measured from the last day of maximal tumescence of successive cycles, and analysis is based on medians because of skew and kurtosis; statistical tests are based on the residuals of an analysis of variance (ANOVA) that first removed the variations between individual females. Both the median and mode of 221 fertile cycles was 38 days (as was the median of 1,240 infertile cycles). Irregularity from this 'norm' was calculated for each cycle as $|\ln(\text{observed}/38)|$, which gives the same deviation for a cycle of 19 days as for 76 days.

Received 15 December 1997; accepted 9 February 1998.

1. vom Saal, F. S., Finch, C. E. & Nelson, J. F. in *The Physiology of Reproduction* 2nd edn (eds Knobil, E. & Neill, J. D.) 1213–1314 (Raven, New York, 1994).
2. Finch, C. E. *Longevity, Senescence and the Genome* (Univ. Chicago Press, 1990).
3. Laws, R. M., Parker, I. S. C. & Johnstone, R. C. B. *Elephants and their Habitats* (Clarendon, Oxford, 1975).
4. Williams, G. C. Pleiotropy, natural selection, and the evolution of senescence. *Evolution* 11, 398–411 (1957).
5. Hamilton, W. D. The moulding of senescence by natural selection. *J. Theor. Biol.* 12, 12–45 (1966).
6. Hawkes, K., O'Connell, J. F. & Blurton-Jones, N. Hadza women's time allocation, offspring provisioning, and the evolution of long postmenopausal life spans. *Curr. Anthropol.* 38, 551–565 (1997).
7. Rogers, A. R. Why menopause? *Evol. Ecol.* 7, 406–420 (1993).
8. Charlesworth, B. *Evolution in Age-structured Populations* (Cambridge Univ. Press, Cambridge, 1980).
9. van Noord, P. A. H., Dubas, J. S., Dorland, M., Boersma, H. & te Velde, E. Age at natural menopause in a population-based screening cohort: the role of menarche, fecundity, and lifestyle factors. *Fertil. Steril.* 68, 95–102 (1997).
10. Gouzoules, S. & Gouzoules, H. in *Primate Societies* (eds Smuts, B. B. et al.) 299–305 (Univ. Chicago Press, 1987).
11. Packer, C., Scheel, D. & Pusey, A. E. Why lions form groups: food is not enough. *Am. Nat.* 136, 1–19 (1990).
12. McComb, K., Packer, C. & Pusey, A. E. Roaring and numerical assessment in contests between groups of female lions, *Panthera leo*. *Anim. Behav.* 47, 379–387.
13. Heinsohn, R. & Packer, C. Complex cooperative strategies in group-territorial African lions. *Science* 269, 1260–1262 (1995).
14. Pusey, A. E. & Packer, C. Non-offspring nursing in social carnivores: minimizing the costs. *Behav. Ecol.* 5, 362–374 (1994).
15. Reznick, D. Costs of reproduction: an evaluation of the empirical evidence. *Oikos* 44, 257–267 (1985).
16. Vaupel, J. W. & Iashin, A. I. Heterogeneity's ruses: some surprising effects of selection on population dynamics. *Am. Stat.* 39, 177–185 (1985).
17. Medawar, P. B. *The Uniqueness of the Individual* (Dover, New York, 1952).
18. Caswell, H. *Matrix Population Models* (Sinauer, Sunderland, MA, 1989).
19. Wise, P. M., Krajnak, K. M. & Kashon, M. L. Menopause: The aging of multiple pacemakers. *Science* 273, 67–70 (1996).
20. Stolwijk, A. M., Zielhuis, G. A., Sauer, M. V., Hamilton, C. J. C. M. & Paulson, R. J. The impact of the woman's age on the success of standard and donor in vitro fertilization. *Fertil. Steril.* 67, 702–710 (1997).
21. Hill, K. & Hurtado, A. M. The evolution of premature reproductive senescence and menopause in human females: an evaluation of the "Grandmother Hypothesis." *Hum. Nature* 2, 313–350 (1991).
22. Packer, C. et al. Reproductive constraints on aggressive competition in female baboons. *Nature* 373, 60–63 (1995).
23. Packer, C. & Pusey, A. E. The Lack clutch in a communal breeder: lion litter size is a mixed evolutionarily stable strategy. *Am. Nat.* 145, 833–841 (1995).
24. Altmann, S. A. The pregnancy sign in savannah baboons. *Lab. Anim. Digest* 6, 7–10 (1970).
25. Lee, E. T. *Statistical Methods for Survival Data Analysis*, 2nd edn (J. Wiley, New York, 1992).

Acknowledgements. We thank E. C. Birney, M. Borgerhoff Mulder, E. L. Charnov, C. E. Finch, K. Hawkes, P. A. Hunt, M. McClintock, D. Promislow, A. R. Rogers, F. S. vom Saal and P. M. Wise for comments and discussion. Fieldwork on the baboons was supported by the Jane Goodall Institute and grants from the Physical Anthropology program at NSF; the lion study was supported by the LTREB program at NSF.

Correspondence and requests for materials should be addressed to C.P. (e-mail: packer@biosci.umn.edu).

Increased auditory cortical representation in musicians

Christo Pantev, Robert Oostenveld, Almut Engelien*, Bernhard Ross, Larry E. Roberts† & Manfred Hoke

Biomagnetism Center, Institute of Experimental Audiology, University of Münster, D-48129 Münster, Germany

* Neurology Department, University of Münster, D-48129 Münster, Germany

† Department of Psychology, McMaster University, Hamilton, Ontario L8S 4K1, Canada

Acoustic stimuli are processed throughout the auditory projection pathway, including the neocortex, by neurons that are aggregated into 'tonotopic' maps according to their specific frequency tunings^{1–3}. Research on animals has shown that tonotopic representations are not statically fixed in the adult organism but can reorganize after damage to the cochlea⁴ or after training the intact subject to discriminate between auditory stimuli⁵. Here we used functional magnetic source imaging (single dipole model) to measure cortical representations in highly skilled musicians. Dipole moments for piano tones, but not for pure tones of similar fundamental frequency (matched in loudness), were found to be enlarged by about 25% in musicians compared with control subjects who had never played an instrument. Enlargement was correlated with the age at which musicians began to practise and did not differ between musicians with absolute or relative pitch. These results, when interpreted with evidence for modified somatosensory representations of the fingering digits in skilled violinists⁶, suggest that use-dependent functional reorganization extends across the sensory cortices to reflect the pattern of sensory input processed by the subject during development of musical skill.

Three groups, comprising right-handed subjects (Edinburgh handedness questionnaire) with no history of otological or neurological disorders and with normal audiological status, were studied. Subjects were fully informed about the experimental procedures and signed a consent form before participation. The first group ($n = 9$) consisted of musical students with absolute pitch and the second group ($n = 11$) of musical students with relative pitch, who had played their instruments for a mean period of 21 ± 6 and 15 ± 3 years, respectively. Musicians in the first and second groups were recruited from the Conservatory in Münster and reported that they practised for an average of 27 ± 14 and 23 ± 12 hours per week, respectively, during the five years preceding the experiment. Musicians who claimed to have absolute pitch were tested before magnetoencephalographic (MEG) measurements. For this purpose a test was developed according to established methods. A randomized sequence of 35 piano tones between H2 and C7 (concert pitch A 440 Hz, American notation) was presented to the subjects. A musician was accepted to have absolute-pitch ability if $>90\%$ of the tones were correctly recognized. Musicians with relative pitch were either self-identified (nine musicians) or did not meet the tested criterion for absolute pitch (two musicians). The third group of subjects (controls; $n = 13$) consisted of students who had never played an instrument. Mean age was 29 ± 6 years for musicians with absolute pitch, 26 ± 5 years for musicians with relative pitch, and 26 ± 4 years for control subjects.

MEG measurements were carried out with a 37-channel BTi Magnes system. From each subject, auditory evoked fields (AEFs) elicited by right-side stimulation were recorded above the left hemisphere. Auditory stimulation (Fig. 1a) consisted of a semirandomized blockwise presentation of the four piano tones C4, C5, C6 and C7 (American notation, having the first harmonics at 262, 523, 1,046 and 2,093 Hz, respectively) and of four pure tones (of 250,

500, 1,000 and 2,000 Hz) that correspond closely with the fundamental frequency of the piano tones. Each tone was presented 128 times while subjects watched cartoon videos intended to fixate their attention. A single equivalent current dipole model was used to explain the field distribution of the auditory evoked field component N1 (occurring ~100 ms after stimulus onset) for each of the eight stimulus conditions. Because the N1 field distributions were dipolar, the applied model was able to explain most of the field variance (goodness of fit calculated over all subjects was within the range of >97%).

In the human auditory cortex, tonotopic representation of the cortical sources corresponding to tones with different spectral content distributes along the medial-lateral axis of the supratemporal plane when measured by the N1 component, with the medial-lateral centre of cortical activation shifting toward the sagittal midline as the frequency of the tone increases⁷. The averaged values of the medial-lateral coordinates of areas of the brain that respond to pure and piano tones, and linear regressions fitted to these data, are shown for the three study groups in Fig. 1b. The dependence of the medial-lateral coordinate on the stimulus frequency was less pronounced for the piano tones; this can be

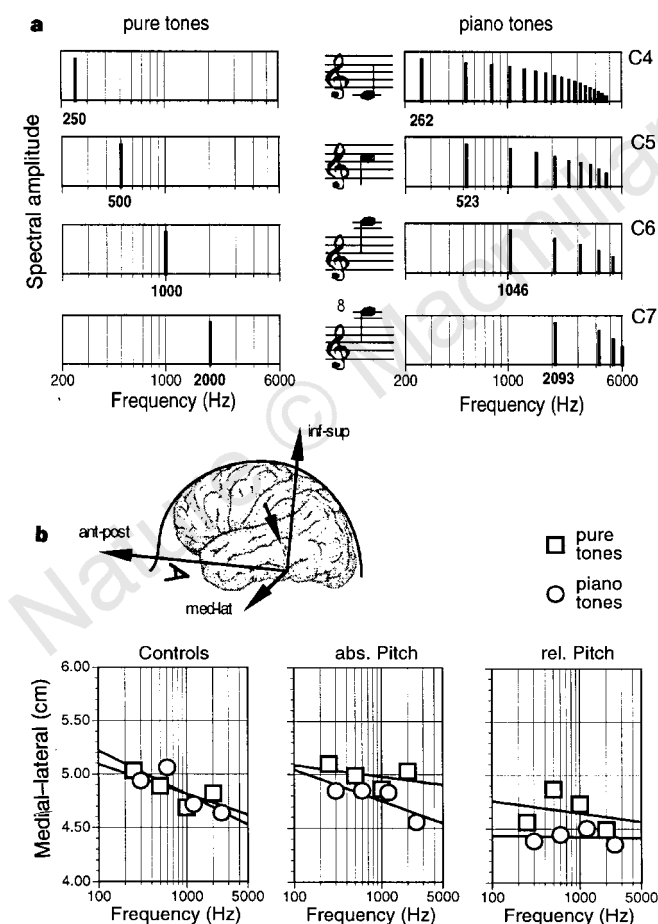


Figure 1 Cortical stimulation by pure and piano tones in musicians and non-musicians. **a**, Spectra of pure tones and piano tones. **b**, Medial-lateral coordinates are shown for single equivalent current dipoles fitted to the field patterns evoked by pure tones and piano tones in control subjects and subjects with absolute or relative pitch. The inset defines the coordinate system of the head. Equivalent current dipoles shift toward the sagittal midline along the medial-lateral coordinate as a function of the frequency of the tone. Ant-post, anterior-posterior; med-lat, medial-lateral; inf-sup, inferior-superior.

explained by the spectral content of these tones. The spatial location of the cortical sources activated by the piano tones is determined by the fundamental frequency of the tones. However, the relative contribution of higher harmonics of the fundamental frequency to the spectral energy is highest for the lowest piano tones and decreases as the fundamental frequency increases (Fig. 1a). No significant differences between the three groups of subjects were seen (Fig. 1b).

A significant difference between the two groups of musicians and the control group was found, however, in the strength of activation of the cortical sources by the piano tones compared with the strength of activation by the pure tones (Fig. 2a). For both groups of musicians with absolute or relative pitch, the equivalent dipole moment was 21–28% larger for piano tones compared with pure tones ($P < 0.0001$, paired *t*-test). This effect was found even though the two types of tone were matched in loudness. For the control group, no significant difference was found in dipole moment in response to piano and pure tones ($P = 0.72$). Given a constant direction of the single-current dipole, the equivalent dipole moment indicates the total strength of cortical activation (the number of neurons simultaneously active at the time of the auditory evoked field N1). Cortical activation may have increased at the time of the N1 event because more neurons were engaged in the processing of these tones in skilled musicians or because the synchrony of neural activity was greater as a consequence of their

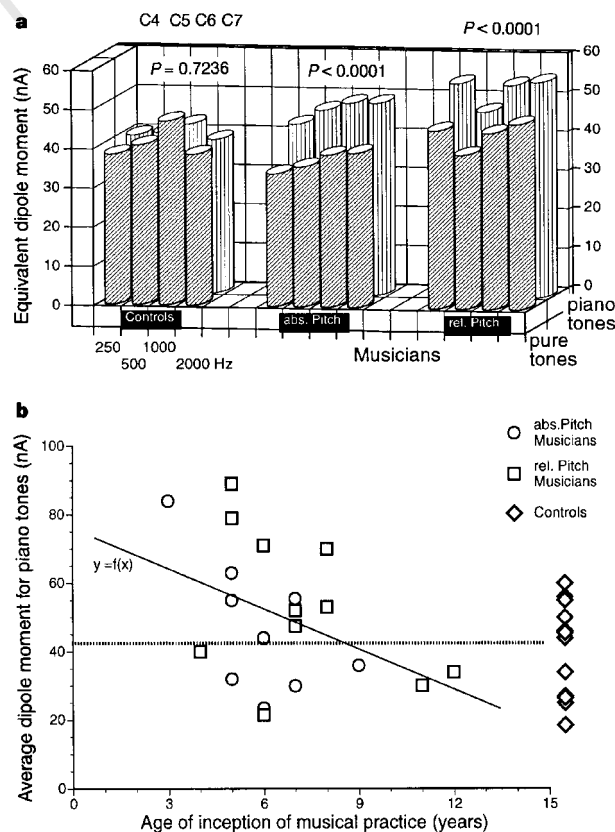


Figure 2 Strength of cortical activation is higher in response to piano, rather than pure, tones and in musicians who began practising before age 9. **a**, Equivalent dipole moment (strength of neuronal activation during the auditorily evoked field N1) is shown for pure tones (diagonally striped columns) and piano tones (vertically striped columns) in control subjects and subjects with absolute or relative pitch. **b**, Mean dipole moment as a function of the age at which musical training began in musicians with absolute pitch or relative pitch. The solid line is the linear correlation fitted to the data of the combined groups ($r = -0.426$, $P = 0.031$). The broken line denotes mean dipole moment in control subjects who never played a musical instrument.

musical training.

Before the experiments, our musicians were interviewed to determine the age at which their musical training had begun. There was a significant linear correlation between mean dipole moment for the piano tones and the age of inception of musical training when the two groups of musicians were combined (coefficient of correlation (r) = -0.426 , $P = 0.037$) (Fig. 2b), and in the absolute-pitch group separately ($r = -0.621$, $P = 0.031$). A similar trend was seen in the relative-pitch group ($r = -0.452$, $P = 0.082$; one-tailed tests). These data indicate that the younger the subjects started playing their instrument, the larger their cortical reorganization in recognition of piano tones was. Enhanced cortical representations were seen in subjects who began to practise before the age of 9 years. This correlation is similar to that reported in a previous study, which examined somatosensory representations of fingering digits in highly skilled string players⁶. In that study, enhancement of the somatosensory representation was most evident in musicians who had begun to practise before the age of 10 years. The similar dependence of auditory and somatosensory representations on practice beginning before the age of about 10 years indicates that cortical reorganization induced by training to learn a musical skill conforms with the pattern of sensory input experienced during practice. Although we cannot separate effects due to duration of training from effects dependent on the age at which training commenced, it seems unlikely that duration of training alone is the critical variable, because our musicians had practised their skill for an average of 18 ± 5 years when our measurements were taken.

The dipole moments evoked by the piano tones did not differ significantly between musicians reporting the piano ($n = 9$) or woodwinds or strings ($n = 11$) as their principal instrument. However, eight of the eleven subjects who reported woodwinds or strings as their principal instrument reported the piano to be a secondary instrument. Therefore, we cannot determine whether cortical reorganization is a specific response to piano tones that were experienced during musical training, or whether it is specialized more broadly for processing of musical or non-musical stimuli with complex harmonic spectra. Enhancement of the cortical representation of piano tones might have been caused by greater attention having been paid to these tones during the experiments by musicians than by non-musicians, despite our video-viewing requirement which was intended to fixate attention, but this seems unlikely. Attentional modulation during testing cannot explain why the degree of cortical activation seen among our skilled musicians depended on the extent of their musical experience gained several years before testing (Fig. 2b), unless the ability to command attention is itself a consequence of cortical reorganization.

The enhanced cortical representation for tones of the musical scale that we observed in musicians corresponds to the results of an earlier magnetic resonance imaging study⁸: a structural enlargement of the planum temporale of the left hemisphere in musicians compared with non-musicians was found. Our data thus associate a use-correlated functional property with cortical architectonics and also raise the possibility that musical experience during childhood may influence structural development of the auditory cortex. A special role for early experience is compatible with evidence for maturation of fibre tracts and the presence of an intracortical neuropil persisting to the age of 7 years (ref. 9). Left-hemispheric enlargement of the planum temporale has been reported to be more pronounced for musicians with absolute pitch than for musicians with relative pitch. The functional reorganization shown here did not relate to this perceptual skill, perhaps because absolute pitch is supported by cortical mechanisms not seen in the auditory N1 response occurring up to 100 ms after tone onset, or is based on temporal rather than structural coding of auditory information. □

Methods

Subjects. Before the experiment we interviewed our musicians to collect information about hours of practice, knowledge of music theory, sight-reading ability, absolute- or relative-pitch ability (people have absolute pitch if they are able to perfectly map specific tone frequencies onto response choices), principal instrument, other instruments played, musicians in the family, and the age at which musical training began. The principal instruments of our musicians were piano (nine subjects), woodwind instruments (seven subjects), and strings (four subjects). Control subjects were asked as part of a standardized pre-experimental interview whether they had any special ability with respect to pitch perception or whether they currently played or had previously played a musical instrument and that to their knowledge they had no special pitch-perception ability. Although the pitch-perception abilities of control subjects were not explicitly tested, the incidence of absolute pitch in the general population is estimated to be $<0.01\%$ (ref. 10). Relative pitch is typically an ability of professional musicians only^{11,12}. All subjects had air-conduction and bone-conduction hearing-level thresholds of no more than 10 dB, in the range from 250 Hz to 8000 Hz. The thresholds for pure and piano tones were measured to within 2 dB for each subject. The intensity of each tone was set at 60 dB above each subject's measured threshold.

Source analysis. Magnetic fields evoked by acoustic stimuli were averaged and filtered within the band 0.1–20 Hz. For each evoked magnetic field a single equivalent current dipole (ECD) was fitted. The means of the dipole moments were computed to within a 40-ms time interval around the maximum of the dipole moment. The coordinates of the dipole location were calculated as a mean of the data points within the time interval between the maximum and the minimum of the evoked field, and satisfied the following requirements: first, a goodness of fit of $>95\%$ of the ECD model to the measured field; second, variation of the source coordinates, around the maximal root mean square field value, of <15 mm within an interval of 40 ms; and third, anatomical distance of the ECD to the midsagittal plane of >2 cm and inferior–superior value of >2 cm. Given a constant direction of the equivalent current dipole, the dipole moment indicates the total strength of cortical activation, that is the number of neurons involved during a cortical response. If this number increases, the dipole moment also increases. Any active focal area can be modelled by an equivalent current dipole. Each dendritic current flow has been estimated¹³ to contribute to the dipole moment according to the formula dipole moment = (conductivity) $\times$ (cross-section of the dendrite) $\times$ (potential difference along the dendrite). If the diameter of an apical dendrite is assumed to be $4 \mu\text{m}$ with an intracellular conductivity of about $0.25/\Omega\text{m}$ and a potential difference of 10 mV, about 30,000 dendrites would be necessary to produce a dipole moment of 10 nA (ref. 13). Only currents flowing tangentially with respect to the surface of the volume conductor (head) are detected. Magnetic fields generated by axonal currents are assumed to be quadrupolar and are neglected in the calculation.

Statistical analysis. We expected on the basis of previous results from studies of skilled string players⁶ and animals¹⁴ that changes in cortical representation are more readily induced by sensory experience in the young brain than in the adult brain. One-tailed tests were therefore accepted for evaluation of age-related effects (negative correlations were predicted). Although it made no difference to the outcome, two-tailed probabilities are reported for group comparisons.

Received 1 August 1997; accepted 20 January 1998.

- Merzenich, M. M., Knight, P. L. & Roth, G. L. Cochleotopic organization of primary auditory cortex in the cat. *Brain Res.* **63**, 343–346 (1973).
- Merzenich, M. M., Kaas, J. H. & Roth, G. L. Comparison of tonotopic maps in animals. *J. Comp. Neurol.* **166**, 387–402 (1976).
- Rauschecker, J. P., Tian, B. & Hauser, M. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* **268**, 111–114 (1995).
- Irvine, D. R. F. & Rajan, R. *Advances in Hearing Research* (eds Manley, G. A., Klump, G. M., Kappl, C., Fastl, M. & Oeckinghaus, H.) 3–23 (World Sci Publ Co, Singapore, 1995).
- Recanzone, G. H., Schreiner, C. E. & Merzenich, M. M. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.* **13**, 87–103 (1993).
- Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B. & Taub, E. Increased cortical representation of the fingers of the left hand in string players. *Science* **270**, 305–307 (1996).
- Pantev, C. et al. Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroencephalogr. Clin. Neurophysiol.* **69**, 160–170 (1988).
- Schlaug, G., Jancke, L., Huang, Y. & Steinmetz, H. *In vivo* evidence of structural brain asymmetry in musicians. *Science* **267**, 699–701 (1995).

9. Yakovlev, P. I. & Lecours, A. in *Regional Development of the Brain in Early Life* (ed. Minkowski, A.) 3–70 (Blackwell, Oxford, 1967).
10. Takeuchi, A. H. & Hulse, S. H. Absolute pitch. *Psychol. Bull.* **113**, 345–361 (1993).
11. Barnea, A., Granot, R. & Pratt, H. Absolute pitch–electrophysiological evidence. *Int. J. Psychophysiol.* **16**, 29–38 (1994).
12. Benguerel, A. P. & Westdal, C. Absolute pitch and the perception of sequential music intervals. *Music Percept.* **9**, 105–120 (1991).
13. Williamson, S. J. & Kaufman, L. in *Auditory Evoked Magnetic Fields and Electric Potentials* (eds Grandori, F., Hoke, M. & Romani, G. L.) 1–39 (Karger, Basel, 1990).
14. Kaas, J. H., Merzenich, M. M. & Killackey, H. P. The reorganization of somatosensory cortex following peripheral nerve damage in adult and developing mammals. *Annu. Rev. Neurosci.* **6**, 325–356 (1983).

Acknowledgements. We thank our subjects for their cooperation, K. Berning for assistance and A. Wollbrink for engineering. This research was supported by grants from the Deutsche Forschungsgemeinschaft and NATO.

Correspondence and requests for materials should be addressed to C.P. (e-mail: pantev@uni-muenster.de).

Dynamic cortical activity in the human brain reveals motor equivalence

J. A. S. Kelso*, A. Fuchs*, R. Lancaster*, T. Holroyd*, D. Cheyne† & H. Weinberg†

* Program in Complex Systems and Brain Sciences, Center for Complex Systems, Florida Atlantic University, Boca Raton, Florida 33431-0991, USA

† Brain Behavior Laboratory, Discovery Park, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

That animals and humans can accomplish the same goal using different effectors and different goals using the same effectors attests to the remarkable flexibility of the central nervous system. This phenomenon has been termed ‘motor equivalence’^{1,2}, an example being the writing of a name with a pencil held between the toes or teeth. The idea of motor equivalence has reappeared because single-cell studies in monkeys have shown that parameters of voluntary movement (such as direction) may be specified in the brain, relegating muscle activation to spinal interneuronal systems^{3,4}. Using a novel experimental paradigm⁵ and a full-head SQUID (for superconducting quantum interference device) array to record magnetic fields corresponding to ongoing brain activity, we demonstrate: (1), a robust relationship between time-dependent activity in sensorimotor cortex and movement velocity, independent of explicit task requirements; and (2) neural activations that are specific to task demands alone. It appears, therefore, that signatures of motor equivalence in humans may be found in dynamic patterns of cortical activity.

The first experiment required human volunteers to flex or extend the preferred index finger either on the beat of a metronome or between metronome beats, the frequency of which was fixed at 1 Hz (see Methods). Notice that these experimental conditions may be grouped either according to the kinematics of motion (flexion versus extension movements and their derivatives) or according to the coordination task (synchronization or syncopation). Figure 1 shows the relative phase between stimuli and movement peaks across cycles for all four conditions. As requested, the peak of the movement is synchronized closely to the stimulus in the flexion-on-the-beat and extension-on-the-beat conditions. Likewise, subjects are able to place a movement between stimuli in the flexion-off-the-beat and extension-off-the-beat, syncopation conditions. Also shown are histograms that measure task success. In general, the distributions for the syncopation conditions are broader than those for synchronization, indicating performance is more variable. Subjects had more difficulty syncopating than synchronizing, which conforms to everyday experience and behavioural evidence⁵.

Brain activity was recorded continuously during these tasks, using a 64-channel magnetometer sampled at 250 Hz. All the magnetic fields generated by the brain arise from the flow of electrical current, which is mainly ionic and originates in the dendrites and cell bodies of cortical neurons⁶. Figure 2 gives an example of the field pattern shown on a model head (Fig. 2a) and the same pattern shown in polar projection on the plane (Fig. 2b). The dominant field over the sensorimotor area of the left hemisphere is expected for right-handed movements. How is this evolving brain activity related to the behaviour produced? Figure 3a shows cortical activity patterns averaged across subjects for each task sampled at various time points (red line) throughout the movement. The average amplitude profile of the movement is also shown (green line). To ease viewing across conditions, the movement profiles are all plotted in the same positive-going direction. Notice the strong dipolar field in the sensorimotor area of the left hemisphere during the first part of the movement, regardless of whether it involves flexion or extension. Notice also that the field reverses just after the peak movement (column 5) and then becomes much weaker and more distributed. Figure 3b plots the time series of the average brain activity obtained from single sensors for the flexion-on (the beat) condition. On the lower left is plotted the movement profile and its derivative, velocity: particularly on the left side of the array, brain activity and movement velocity are nearly superimposed, as confirmed by correlation analysis. The cross-correlation between movement velocity and time-varying brain signal is almost 1 (or –1) in the yellow area of Fig. 3b, corresponding precisely to maxima and minima of the field.

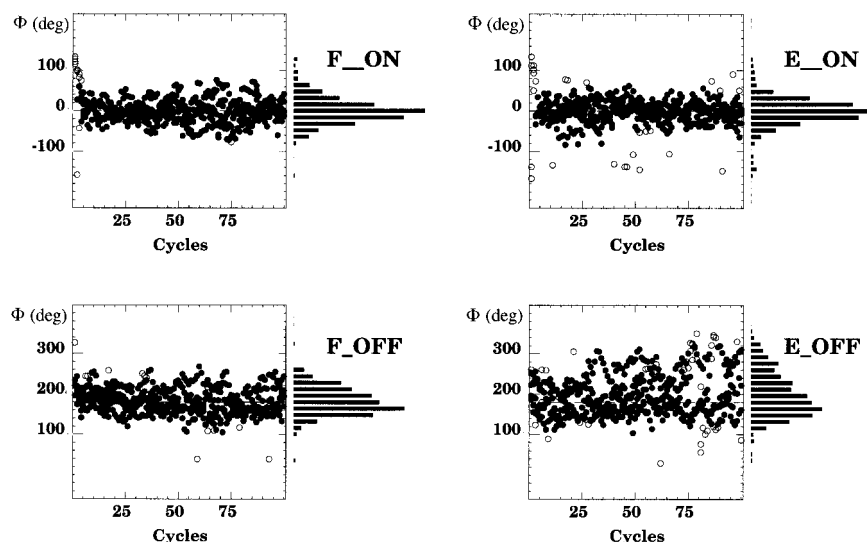


Figure 1 Relative phase across cycles for all conditions (top row, synchronization; bottom row, syncopation) and all subjects, a total of ~500 observations per condition. Solid circles indicate cycles within a ± 60 degree range of the average phase for single subjects. Open circles are outside this range and were rejected from further analysis.

Elsewhere the correlation is much weaker and, in certain anterior and posterior regions, practically zero. This strong correlation between movement velocity and the time course of activity in left sensorimotor cortex is found in all four experimental conditions.

Decomposition of the brain's magnetic field into components corresponding to localized current sources is an ill-posed problem. Nevertheless, the spatial patterns of cortical activity shown in Fig. 3a are well localized and mostly stationary during particular phases of

the task. We decomposed the brain signals into spatial patterns and time-varying amplitudes using two methods. The first, known as Karhunen–Loève (K–L) decomposition, or principal components analysis⁷, calculates the eigenvectors of the covariance matrix between sensors x and y as an orthogonal basis. Tangential currents naturally produce spatial correlations (the field entering the scalp at one location and leaving at another; Figs 2, 3), and the resulting principal components may capture this dipolar structure. The

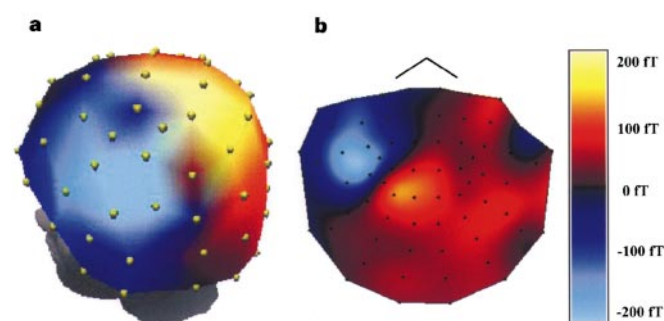


Figure 2 Neuromagnetic activity of the human brain. **a**, Sensor locations displayed on the subject's head, and **b**, corresponding polar projection. Blue indicates sites where the magnetic field is entering the head, and red-yellow indicates sites where the field is leaving the head. The field amplitude is in units of femtoTesla.

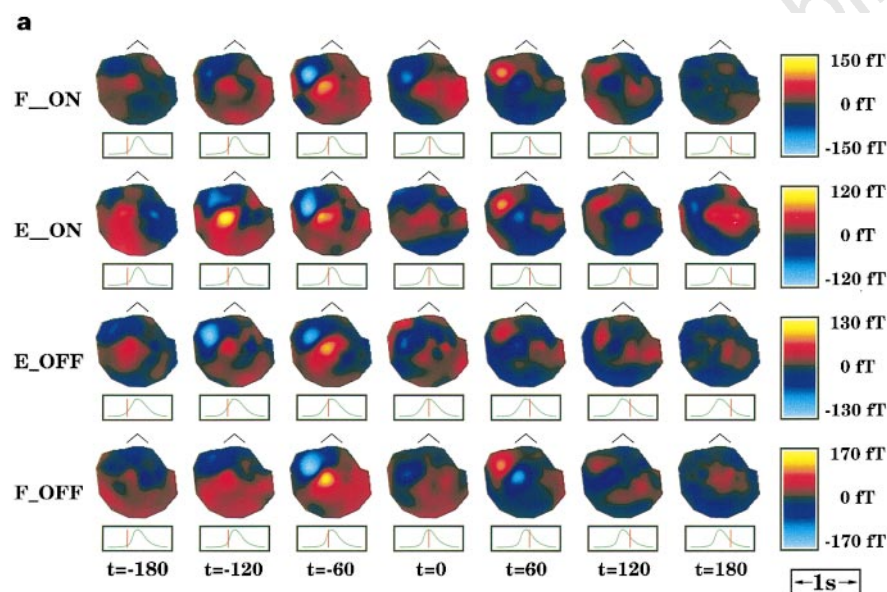
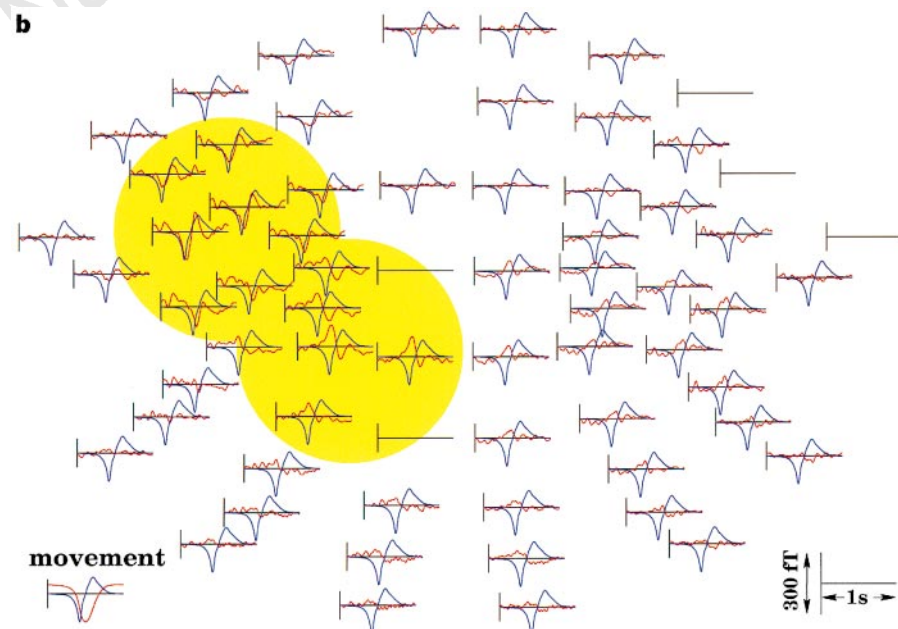


Figure 3 Cortical activity patterns for each experimental condition. **a**, Data are averaged across cycles and subjects at time points t (ms). Time point $t = 0$ is the peak movement amplitude. The green curve in the box shown below each pattern is the average amplitude profile for each condition. The red line indicates the time at which brain activity is sampled during the movement cycle. Each condition is scaled separately to highlight the peak fields. In general, the activity for flexion conditions has a slightly higher amplitude range than the extension conditions. **b**, Time series of averaged brain activity (red) across cycles and subjects for the flexion-on-the-beat condition. Overlaid (blue) is the movement velocity. In the centres of the yellow circles, the correlation between brain activity and movement velocity is close to ± 1 .



group averages shown in Fig. 4a indicate that the first spatial mode captures about 60% of the variance in the brain signals (the eigenvalue for each condition is given in the boxes). Because the top spatial mode, like the underlying neural generators, is under no orthogonality constraint, it is identical to the dominating spatial

pattern of brain activity observed experimentally. Figure 4a shows that its time-dependent amplitude tracks the velocity profile extremely well, especially for the initial velocity peak associated with the active phase of the coordination task (see Methods). The second velocity peak constitutes the less active phase of the task, and the

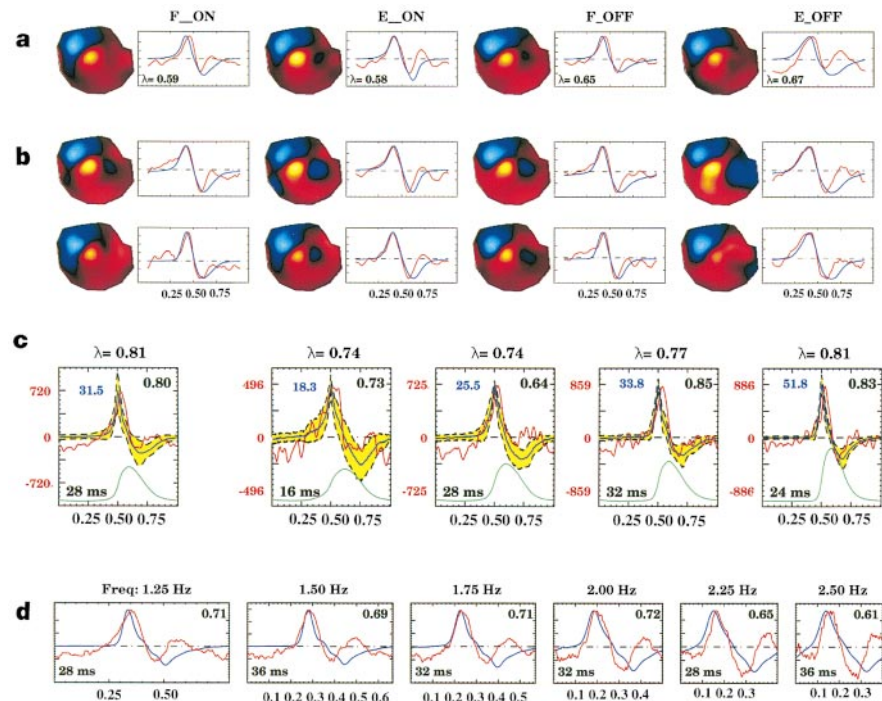


Figure 4 Relationship between cortical activity and movement velocity. **a**, The first spatial mode of a Karhunen–Loève decomposition and its time-dependent amplitude (red) averaged across subjects. Eigenvalues in the lower left corner of each box indicate how much of the variance of the entire signal is contained in a given mode. Movement velocity for each condition is shown in blue. **b**, The spatial pattern that best fits the time series of the movement velocity for the group (first row) and a single subject (second row) for all conditions. **c**, Scaling between brain activity (red) and movement velocity (blue/grey). Left box, average data for all

conditions for the subject shown in **b**. Remaining boxes, data sorted in bins according to different peak velocities (blue numbers), which increase from left to right (standard deviation in yellow, bounded by dashed lines). Amplitude profiles shown in green are on a common scale. Black numbers indicate the eigenvalues (above each box), correlation values between brain activity and velocity profile (top right) and corresponding time lags in ms (bottom left). **d**, Effect of different movement rates. The smaller boxes as one moves from left to right reflect the number of points sampled in a given cycle, which decreases with movement rate.

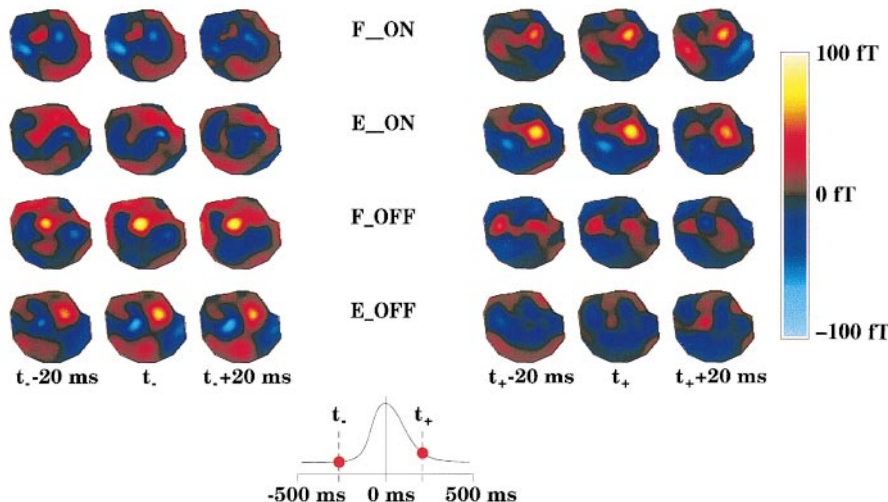


Figure 5 Brain activity fields for the four conditions in the first experiment, flexion and extension movements at 1 Hz, after the dominant spatial pattern (shown in Fig. 4b) was removed from the data set. The residual patterns sampled before the peak movement (left columns) indicate strong similarities across syncopation task conditions but large differences between synchronization and syncopation task conditions. Likewise, the residual patterns sampled after the peak move-

ment (right columns) are similar across synchronization conditions, but different between syncopation and synchronization. The curve at the bottom represents the average amplitude profile for all conditions. Red circles indicate the time points around which brain activity was sampled. t_- and t_+ , Peak field amplitudes for syncopation and synchronization, respectively.

match to the brain signal is weaker. These results are striking because the K-L decomposition simply minimizes the mean-square error without regard to the movement at all. A second method calculates the spatial patterns that best fit the movement and its derivatives⁸. The results of this procedure are shown in Fig. 4b for the group (first row) and a single subject (second row). Once again, the largest mode corresponds to movement-related brain activity (see also Fig. 3a), specifically the velocity of finger flexion or extension.

Two questions arise from these results. First, is the strong relationship between average movement velocity and the time course of cortical activity a coincidence, or is it maintained across changes in the velocity profile? In particular, is this relationship consistent across different peak velocities (with the same length time course) and is it consistent across different movement rates (similar peak velocity, but shorter time course)? Second, what is the influence of the task alone on patterns of brain activation? To answer the first question, we used two approaches. First, we sorted the existing data from all four original experimental conditions into sets representing different peak velocity ranges, from slowest to fastest. The data from four different peak velocity ranges for a representative subject are shown in Fig. 4c. The brain signal (again the time-dependent amplitude of the top K-L spatial mode) is plotted (red) along with the velocity profiles (blue/grey) for the overall data (left box) and four non-overlapping bins, in which velocity increases from left to right. The steepness of the displacement profile (green) reflects the derivative (actual values in blue). Notice that each graph in Fig. 4c is scaled individually, highlighting the correspondence between brain signal and velocity profile. The degree of covariance, given by the correlation value on the top right of each box, is high.

Next, we asked the same subjects to perform the two basic syncopation tasks at six different movement rates (see Methods). Beginning either in the flexion- or extension-off-the-beat conditions, subjects were instructed to syncopate with the metronome, as its rate was increased in small steps of 0.25 Hz, every 10 cycles. Transitions from syncopation to synchronization occur spontaneously in both brain activity and behaviour as movement rate is increased beyond a critical value⁹, but the relationship between velocity and cortical activity has not been examined before to our knowledge. The results were unequivocal across subjects, initial conditions and movement rates. Figure 4d shows a representative example for the flexion-off condition. The same basic dipolar-like spatial pattern was observed at all movement rates. For ease of viewing, the velocity profile (blue) is superimposed on the brain signal (here the best-fit spatial pattern, as in Fig. 4b, though the results for the top K-L mode are nearly identical). The value of the correlation function at the time lag between the first velocity peak and the brain signal is also reported (top right). Given the predictive, rhythmical nature of the task and the probable contribution of reafferent activity from the muscles to sensorimotor cortex^{10–12}, it is unsurprising that the peak movement velocity leads the cortical activity by a small amount (lower left corner of Fig. 4c and d). Yet the similarity in the neural and velocity profiles again is striking, especially in the initial active phase of voluntary movement.

To answer the second question, about task-specific brain activity, we removed the dominant velocity-related spatial pattern from the recorded brain signals. These residual brain patterns are shown in Fig. 5. Notice that the time course of brain activation now is very different for synchronization (top rows) and syncopation (bottom rows) tasks. Before the peak movement (left columns), the brain patterns for syncopation are similar to each other but different from the corresponding brain activity patterns for synchronization. Likewise, after the peak movement (right columns), synchronization conditions are nearly identical and very different from their syncopation counterparts. Any differences attributable to movement direction (for example, flexion-on and extension-off are kinematically similar) are far outweighed by differences due to the task. Thus,

although their detailed physiological interpretation awaits clarification, the patterns of brain activation distinguishing the two tasks are compelling and seem to reflect higher-order task planning.

Finding a coherent neural signal that is specifically related to behaviour in such a complex system as the human brain is a challenge. Although obtained in different experimental circumstances, the discovery of a direct and robust relation between the time course of cortical activity and the movement derivative across a broad range of initial conditions, peak velocities and movement rates fits single-cell studies in monkeys that suggest that speed (in addition to direction^{13,14}) is represented in the discharge rate of motor cortical cells¹⁵. Moreover, our findings help resolve a long-standing question in psychophysical studies of human sensorimotor coordination^{16,17,18}, how the brain coordinates actions in time with external events. It does so by generating a neural activity pattern that significantly mirrors the movement derivative or velocity. At the same time, dynamic patterns of cortical activity are task specific. □

Methods

Task. In the first experiment, subjects ($n = 5$) were instructed to time the peak of their finger movements (synchronize or syncopate) to a visual metronome, which was a computer-controlled light-emitting diode (LED) located in the centre of the visual field, 2 m from the subject. This stimulus was delivered for 100 cycles at a frequency of 1 Hz. For each subject, we randomized the order of the four experimental conditions: flexion on the beat, extension on the beat, flexion off the beat and extension off the beat. A pair of small inflated plastic pillows connected to pressure sensors by plastic tubing was positioned so that the subject's preferred (right) index finger could slide between them. Small-amplitude, almost isometric flexion or extension movements about the metacarpophalangeal joint in the transverse plane increased the pressure in the pillow in the 'active' phase of the task. The passive action of the pillow helped return the finger to its starting position. All motions were transduced by pressure sensors into voltages. The locations of the movement peaks were calculated as the corresponding zero crossings in the derivative of the signal. From these movement peaks stimulus onsets, the relative phase between stimulus and response was calculated on every cycle. In a follow-up experiment that examined the effect of movement rate, the same subjects performed the two syncopation conditions at different metronome frequencies. Beginning in either flexion- or extension-off-the-beat at 1.25 Hz, the stimulus frequency was increased every 10 cycles in steps of 0.25 Hz up to 2.5 Hz. Each subject performed 40 runs in each of the two conditions, a total of 4,800 cycles of movement per subject.

Neuromagnetic measurements. The metronome and pressure-sensor voltages from the response pillows were digitized, along with 64 channels of magnetic field data. The SQUID sensor array (CTF Systems) consisted of 64 first-order gradiometers uniformly distributed over both hemispheres. Each gradiometer had a baseline of 5 cm and a coil diameter of 2 cm. Third gradient response was calculated using a system of reference coils. The sampling rate for all signals was 250 Hz with a 40 Hz low-pass filter and notch filters at 60, 120 and 180 Hz.

Analysis. Two different but related methods of spatiotemporal analysis were applied to the dataset. One was Karhunen–Loève decomposition or principal components analysis, in which the covariance matrix of the time series for single sensors is calculated and diagonalized. The eigenvectors of this matrix represent spatial patterns that are used as basis vectors for the decomposition of the signal. How much power these single patterns contribute to the entire signal is given by the eigenvalue corresponding to the pattern. In many cases, only a few of these modes (here two or three) are needed to capture 80–90% of the variance. A time series (Fig. 4) for each of the spatial patterns is calculated by projecting the signal onto the basis vectors. This method simply creates an orthogonal basis with the axes pointing in the direction of maximal variance (for the higher modes, under the orthogonality constraint), although the dominating mode is related closely to the movement velocity (Fig. 4a). Therefore, we applied a second method that actually takes the velocity of movement into account, calculating a spatial pattern that best fits a given time series. The aim of this method is to find a spatial pattern that yields a minimum

mean-square error if multiplied with a given time series (say, movement velocity) and subtracted from the original signal at each time point. This pattern is given by $P_i = \langle v(t)H_i(t) \rangle / \langle v^2(t) \rangle$, where P_i is the pattern at the location of the i th sensor, $H_i(t)$ is the activity in this sensor, $v(t)$ is the time series (here, movement velocity) and $\langle \dots \rangle$ denotes a temporal average.

Received 16 September 1997; accepted 30 January 1998.

1. Lashley, K. S. Basic neural mechanisms in behavior. *Psychol. Rev.* **37**, 1–24 (1930).
2. Hebb, D. O. *The Organization of Behavior: A Neuropsychological Theory* (Wiley, New York, 1949).
3. Bizzi, E., Mussa-Ivaldi, F. & Giszter, S. Computations underlying the execution of movement: a biological perspective. *Science* **253**, 287–291 (1991).
4. Georgopoulos, A. P. Neural networks and motor control. *Neuroscientist* **3**, 52–60 (1997).
5. Kelso, J. A. S., DelColle, J. D. & Schönner, G. in *Attention and Performance XIII* (ed. Jeannerod, M.) 139–169 (Erlbaum, Hillsdale, New Jersey, 1990).
6. Romani, G. L. in *Advances in Biomagnetism* (eds Williamson, S. J., Hoke, M., Stroink, G. & Kotani, M.) 33–46 (Plenum, New York, 1989).
7. Fuchs, A., Kelso, J. A. S. & Haken, H. Phase transitions in the human brain: spatial mode dynamics. *Int. J. Bifurc. Chaos* **2**, 917–939 (1992).
8. Uhl, C., Friedrich, R. & Haken, H. Analysis of spatiotemporal signals of complex systems. *Phys. Rev. E* **51**, 3890–3900 (1995).
9. Kelso, J. A. S. et al. A phase transition in human brain and behavior. *Phys. Lett. A* **169**, 134–144 (1992).
10. Cheyne, D. & Weinberg, H. Neuromagnetic fields accompanying unilateral finger movements: pre-movement and movement-evoked fields. *Exp. Brain Res.* **78**, 604–612 (1989).
11. Kristeva, R., Cheyne, D. & Deecke, L. Neuromagnetic fields accompanying unilateral and bilateral voluntary movements: topography and analysis of cortical sources. *Electroencephalogr. Clin. Neurophysiol.* **81**, 284–298 (1991).
12. Wiesendanger, M. & Miles, T. S. Ascending pathway of low-threshold muscle afferents to the cerebral cortex and its possible role in motor control. *Physiol. Rev.* **62**, 1234–1270 (1982).
13. Georgopoulos, A. P., Kalaska, J. F., Caminiti, R. & Massey, J. T. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci.* **2**, 1527–1537 (1982).
14. Schwartz, A. B., Kettner, R. E. & Georgopoulos, A. P. Primate motor cortex and free arm movements to visual targets in three dimensional space. I. Relations between cell discharge and direction of movement. *J. Neurosci.* **8**, 2913–2927 (1988).
15. Moran, D. W. & Schwartz, A. B. Motor cortical representation of speed and direction during reaching. *J. Neurophysiol.* (in the press).
16. Bartlett, N. R. & Bartlett, S. C. Synchronization of a motor response with an anticipated sensory event. *Psychol. Rev.* **66**, 203–218 (1959).
17. Wing, A. M. & Kristofferson, A. B. The timing of interresponse intervals. *Percept. Psychophys.* **12**, 455–460 (1973).
18. Dunlap, K. Reactions to rhythmic stimuli. *Psychol. Rev.* **17**, 399–416 (1910).

Acknowledgements. This research was supported by NIMH (U.S.A.), SERC (Canada) and the Human Frontier Sciences Program. We thank M. Burbank and J. Vrba for use of facilities at Port Coquitlam, British Columbia.

Correspondence and requests for materials should be addressed to J.A.S.K. (e-mail: kelso@walt.ccs.fau.edu). For data animation (MPEG format), see <http://www.ccs.fau.edu>.

Lipids are required for directional pollen-tube growth

Mieke Wolters-Arts, W. Mary Lush* & Celestina Mariani

Graduate School Experimental Plant Sciences, Department of Experimental Botany, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

* Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Parkville, Victoria 3052, Australia

Successful pollination and fertilization are absolute requirements for sexual reproduction in higher plants. Pollen hydration, germination and penetration of the stigma by pollen tubes are influenced by the exudate on wet stigmas¹ and by the pollen coat in species with dry stigmas^{2–5}. The exudate allows pollen tubes to grow directly into the stigma, whereas the pollen coat establishes the contact with the stigma. Pollen tubes then grow into the papillae, which are covered by a cuticle. The components of the exudate or pollen coat that are responsible for pollen tube penetration are not known. To discover the role of the exudate, we tested selected compounds for their ability to act as functional substitutes for exudate in the initial stages of pollen-tube growth on transgenic stigmaless tobacco plants¹ that did not produce exudate. Here we show that lipids are the essential factor needed for pollen tubes to penetrate the stigma, and that, in the presence of these lipids, pollen tubes will also penetrate leaves. We propose

that lipids direct pollen-tube growth by controlling the flow of water to pollen in species with dry and wet stigmas.

In transgenic, female-sterile tobacco plants in which the secretory zone of the stigma is ablated by expression of the cytotoxic gene *STIG 1-barnase*, pollen tubes only penetrate the stigma and successfully fertilize ovules when stigma exudate of wild-type plants is applied to the ablated stigmas^{1,6}. To determine whether stigma function is also restored by exudates of other plant species, we applied the lipid-rich⁷ exudate of *Petunia* (a close relative of tobacco), and the carbohydrate-rich⁸ exudate of lily (a distantly related species) to the stigmas. With *Petunia* exudate, pollen tube penetration and seed setting occurred (Fig. 1a). Application of lily exudate resulted in pollen hydration and germination, but not penetration (Fig. 1b), indicating that lipids in the exudate may play a role in pollen tube penetration.

The lipid fraction of the exudate contains many saturated and unsaturated fatty acids, which usually occur as triacylglycerides^{9,10}. To determine whether triacylglycerides are involved in pollen tube penetration, we applied a purified oil with a fatty-acid composition similar to that of exudate to stigmaless pistils. Application of this oil restored stigma function and resulted in seed production. Addition of the same oil in a 1:1 mixture of oil and lily exudate also restored the ability of pollen tubes to penetrate the stigma (see Methods). To investigate the effect of individual components of the purified oil, we applied the unsaturated triacylglycerides triolein, trilinolein and trilinolenin. Only application of trilinolenin allowed penetration of enough pollen tubes into the style (Fig. 1c) to result in seed production. In contrast, a liquid mixture of saturated triacylglycerides did not result in pollen tube penetration (Fig. 1d; see Methods).

We investigated the effects of the products of breakdown of triacylglycerides by applying glycerol, free fatty acids (oleic, linoleic or linolenic acid), monolinolein or dilinolein. None of these products allowed penetration of enough pollen tubes into the style (see Methods). In the presence of trilinolein, the *trans* form of trilinolenin, pollen tubes penetrated the stylar tissue, but were immediately arrested and exhibited severe deformations. Oils that are not based on triacylglycerides, such as silicon oils, did not restore stigma function (see Methods). These data indicate that *cis*-unsaturated triacylglycerides are sufficient and essential for the normal functioning of the exudate.

In genera with dry stigmas, such as *Brassica* and *Arabidopsis*, the

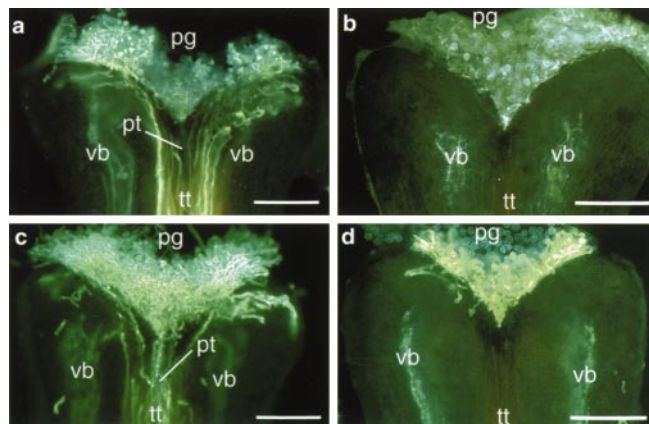


Figure 1 Fluorescence micrographs of longitudinal sections of stigmaless pistils 24 h after pollination. **a**, After application of *Petunia* exudate, pollen grains on the stigmaless surface become hydrated and germinate and the pollen tubes penetrate the stylar tissue. The effects of application of **b**, *Lilium* exudate, **c**, trilinolenin, and **d**, saturated triacylglycerides are also shown (see Methods). Bars represent 100 μ m. Abbreviations are as follows: pg, pollen grain; pt, pollen tube; tt, transmitting tissue; vb, vascular bundle.

pollen coat flows out from the exine to form a 'foot' (contact zone) between the grain and the stigma surface^{11,12}. The pollen coat contains lipids that are required for pollen hydration^{2,3,13,14}. It has been suggested that, during the evolution of dry stigmas, the pollen coat assumed the role of the exudate of wet stigmas^{15,16}. Our transgenic tobacco plants have lost the stigma's secretory zone and thereby the capacity to secrete exudate¹, but have retained the presence of a thin cuticle¹⁷ (Fig. 2c). Therefore, one could infer that stigmaless pistils may function as dry stigmas. Pollen of *Brassica napus*, *B. oleracea* and *Arabidopsis thaliana*, when placed on the ablated stigmas, formed a foot (Fig. 2a,b), and pollen tubes penetrated the pistil, but they were arrested at a length of $\sim 100\ \mu\text{m}$ (Fig. 2c). This supports the view that the pollen coat is functionally equivalent to the exudate, and suggests that lipids may also be required for pollen-tube penetration in species with dry stigmas.

The *Arabidopsis* mutants *cer1*, *cer3*, *cer6* and *pop1* are defective in synthesis of long-chain lipids and their pollen does not become hydrated on stigmas^{2,18} (Fig. 2d). Application of trilinolein enabled mutant pollen grains to produce tubes that penetrated the stigma

(Fig. 2e) and silique developed. Thus, triacylglycerides restore hydraulic contact between pollen grain and stigma in these mutants, allowing pollen to become hydrated.

The flower is the only part of the plant in which fluid or semifluid lipids are secreted at maturity on external surfaces, such as the stigma and pollen exine^{9,14,19}. To determine whether exudate can replace the stigmatoid tissue, we removed the cuticle of a mature leaf, applied *Petunia* exudate or trilinolein and added pollen. Tobacco pollen applied alone to the stripped leaf surface did not germinate, but, in the presence of exudate or trilinolein, pollen tubes penetrated the intercellular spaces of the parenchyma cells of the leaf and grew towards the abaxial side (Fig. 3a). Application of *Brassica* pollen grains (in the absence of exudate or oil) also resulted in penetration of the leaf by pollen tubes (Fig. 3b). No penetration occurred when pollen-germination medium, saturated triacylglycerides, dilinolein or silicon oils were applied to the leaf surface. Germination does not occur on a mature leaf when the cuticle is still present. However, pollen grains of tobacco in the presence of trilinolein or pollen grains of *Brassica* (in the absence of oil) were able to germinate and penetrate a young leaf with a thin cuticle (data

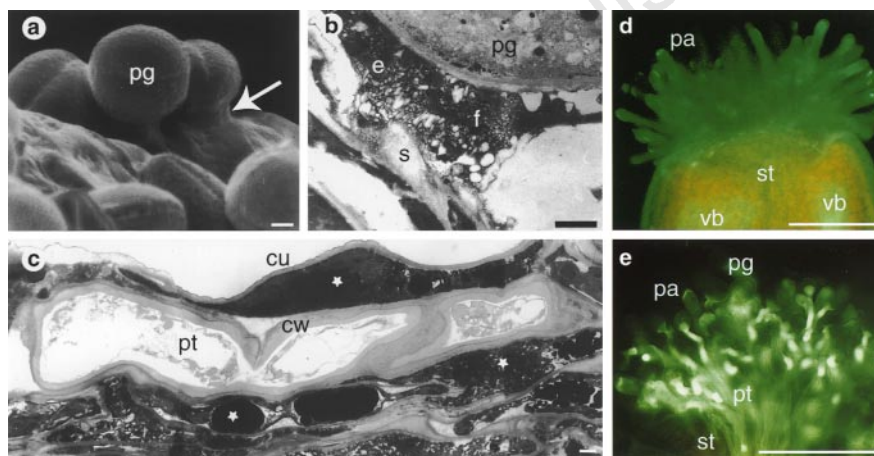


Figure 2 Growth of *Brassica* pollen on stigmaless pistils, and lack of growth of *Arabidopsis pop1* mutant pollen after self-pollination of *Arabidopsis*. **a-c**, Germination and tube growth of *Brassica* pollen on stigmaless pistils 24 h after pollination. **a**, SEM image showing the foot between the pollen grain and the stigmaless surface (arrow). **b**, TEM image showing the electron-opaque appearance in the foot characteristic of compatible pollen-stigma interactions. **c**, Pollen tubes penetrated the stigmaless pistil and grew between the

compressed, degenerated cells (asterisk) of the ablated stigma. Bar represents $1\ \mu\text{m}$. **d,e**, Effect of lipids on *Arabidopsis pop1* mutants 24 h after self-pollination. **d**, Pollen grains did not adhere to the papillae and were washed away during the preparation. **e**, After application of trilinolein to the stigma, pollen tubes penetrated the style. Bars represent $50\ \mu\text{m}$. Abbreviations are as follows: cu, cuticle; cw, cell wall; e, exine; f, foot; pa, papilla; pg, pollen grain; pt, pollen tube; s, surface; st, style; vb, vascular bundle.

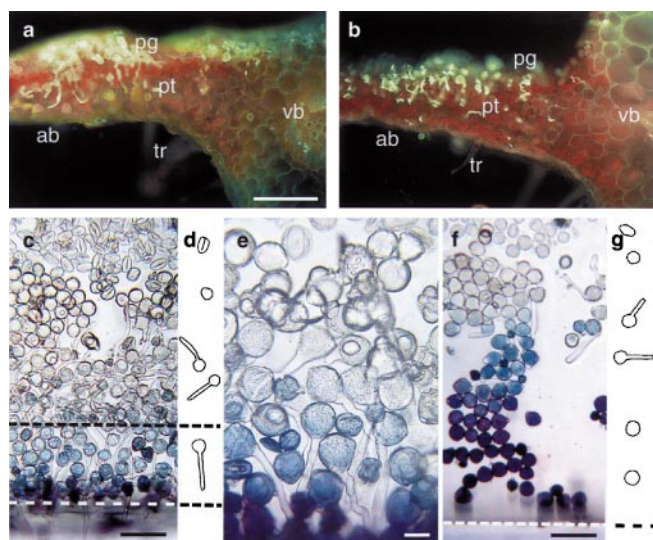


Figure 3 Growth of pollen on leaves and in oil bordering gelled germination medium. **a,b**, Germination and tube growth of pollen on the abaxial side of tobacco leaves from which the cuticle was removed (transverse sections, 24 h after pollen application). **a**, When trilinolein was applied to leaves, tobacco pollen became hydrated and germinated, and pollen tubes penetrated into the leaf. **b**, *Brassica* pollen applied to leaves also germinated and tubes penetrated the tissue. Bar represents $100\ \mu\text{m}$. **c-g**, Germination and tube growth of tobacco pollen in oil bordering gelled germination medium (3 h after establishment of cultures). The gradient of blue indicates hydration of the pollen (toluidine blue was incorporated in the medium). **c**, The pollen grains in trilinolein that are nearest to the gel (white dashed line) produce tubes that grew towards and into the agarose. Further away from the boundary, growth is non-directional (black dashed line); some pollen grains became hydrated but did not germinate and the furthest away did not hydrate. Bar represents $5\ \mu\text{m}$. **d**, Diagrammatic representation of the events shown in **c**. **e**, Close-up view of **c**. Bars represent $10\ \mu\text{m}$. **f**, Pollen grains in saturated triacylglycerides. Bar represents $5\ \mu\text{m}$. **g**, Diagrammatic representation of the events shown in **f**. Abbreviations are as follows: ab, abaxial; pg, pollen grain; pt, pollen tube; tr, trichome; vb, vascular bundle.

not shown). This indicates that the state of the cuticle is also crucial for pollen tube penetration^{12,20,21}. Our results show that penetration of tissues by pollen tubes depends on the presence of lipid carried either by the coat on the pollen of *Brassica* or by the exudate, and it can occur regardless of the cell type that the pollen encounters.

We have shown that *cis*-unsaturated triacylglycerides and the pollen coat enable pollen-tube growth within specialized (dry or wet stigmas) or unspecialized (leaf) tissues. The hydraulic contact between pollen grain and the underlying tissue is established by the pollen coat or exudate^{2,3,5,7}. To investigate the role of triacylglycerides in water transfer between the underlying tissue and the pollen, we created artificial boundaries between oil and a pollen-germination medium. Pollen grains were placed in the oil and a gelled germination medium containing toluidine blue replaced the plant tissue. In *cis*-unsaturated triacylglycerides, pollen grains nearest to the interface became hydrated and germinated, and the pollen tubes grew towards and into the agarose (Fig. 3c–e). In contrast, in saturated triacylglycerides, fewer pollen grains germinated and their tubes did not grow directionally (Fig. 3f, g). Similar results were obtained when water replaced the germination medium in the gel. Differences in the distribution of toluidine blue stain indicate a difference in water supply when different oils were used. The results show that the two oils have different effects on pollen germination and pollen-tube growth.

To confirm that directional growth was a response to a directional supply of water, we disordered the oil phase by replacing it with an emulsion of unsaturated triacylglycerides and germination medium (ratio 1:1). Pollen-tube growth was no longer directional, and tobacco pollen tubes did not penetrate wild-type stigmas bearing exudate when the grains were pre-imbibed in germination medium (data not shown). Thus, directed pollen-tube growth *in vivo* and *in vitro* occurs only when there is a directional supply of water. The control that lipids exert on the movement of water depends on the physicochemical characteristics of the lipids. This control may be a direct consequence of the properties of the lipid as a matrix in which water is distributed²², or an indirect consequence resulting from effects of unsaturated lipids on the permeability of the plasma membrane of the pollen and/or stigmatic cells. The importance of controlled water flow concurs with findings that this flow plays a crucial role in pollen–pistil interactions^{5,23,24}.

Cis-unsaturated triacylglycerides are therefore required for penetration of the stigma by tobacco pollen tubes. These lipids can functionally replace both the exudate and the stigmatoid tissue, and even allow pollen tubes to penetrate leaf tissue. The pollen coat of species with dry stigmas has the same function as the exudate. We propose that the lipids present during pollen–stigma interactions are necessary to regulate water uptake by the pollen grain/tube, establishing an external water gradient. The strength of the gradient, which in pollen–stigma interactions is determined by the composition of the lipids, steers directional growth^{25–30}. □

Methods

Plant material. Transgenic *STIG 1-barnase* tobacco plants¹, *Nicotiana tabacum* cv. Petit Havana SR1, *Petunia hybrida* W166K, *Arabidopsis thaliana*, *Arabidopsis eceriferum* *cer1*, *cer3*, *cer6* and *pop1* mutant plants, *Brassica napus*, *Brassica oleracea* and *Lilium longiflorum* cv. Snow Queen were grown under growth-chamber conditions (day phase: 15 h 20 °C, 65% relative humidity; dark phase: 9 h 18 °C, 65% relative humidity).

Studies of pollen-tube growth *in vivo*. Flowers of transgenic *STIG 1-barnase* tobacco plants were emasculated before anthesis. Exudate or 2 µl selected medium was applied to the stigmaless surface before it was pollinated with excess fresh pollen. In general, when ≥15 pollen tubes penetrated the style, seed production occurred.

Applied compounds. Germination medium was 10% sucrose and 0.01% boric acid. The different oil and lipid compounds used were as follows: chromatographically purified olive oil (fatty acids 10% palmitic acid (16:0), 2% stearic acid (18:0), 83% oleic acid (18:1, [*cis*]-9), 4% linoleic acid (18:2, [*cis*]-

9,12) and 1% linolenic acid (18:3, [*cis,cis,cis*]-9,12,15)) (Larodan, Sweden); the triacylglycerides triolein (C18:1, [*cis*]-9), trilinolein (18:2, [*cis,cis*]-9,12), trilinolenin (18:3, [*cis,cis,cis*]-9,12,15) and trilinolelaidin (C18:2, [*trans,trans*]-9,12) (Sigma); a mix of saturated triacylglycerides (equal parts by mass of trihexanoin (C6:0), triheptanoin (C7:0), trioctanoin (C8:0), trinonanoin (C9:0) and tridecanoin (C10:0)) (Larodan, Sweden); 1,3 dilinolein (18:2, [*cis,cis*]-9,12) (Sigma); 1-monolinolein (18:2, [*cis,cis*]-9,12) (Sigma); the fatty acids oleic acid (C18:1, [*cis*]-9), linoleic acid (C18:2, [*cis,cis*]-9,12) and linolenic acid (18:3, [*cis,cis,cis*]-9,12,15) (Sigma); glycerol (Sigma); and the silicon oils AR20, AR200 (phenylmethyl-polysiloxane), CR350 (chlorphenylmethyl siloxane) (Wacker-Chemie, Germany), DC200 (60,000 mm² s⁻¹ dimethyl siloxane polymer) and DC344 (polydimethyl cyclosiloxane tetramer) (Dow Corning Europe, Belgium).

24 h after application of compounds and pollination, we collected pistils. Vibratome-cut sections of ~150 µm were placed in a drop of aniline blue stain (0.1% in 0.1M K₃PO₄) and examined with a Leitz Wetzlar orthoplan microscope equipped for ultraviolet epifluorescence, using the appropriate filters. Hand-cut sections of pollinated leaves and pistils of *Arabidopsis* were pretreated in 2M NaOH at 60 °C for 10 min to soften the tissue. Pistil squashes were used to count the number of pollen tubes that had penetrated the stigma. Pistils were fixed in ethanol/acetic acid (3:1) for 12 h or more at 4 °C and in 2M NaOH for 1 h at 60 °C, and were then rinsed with water and placed directly into a drop of aniline blue stain.

Statistical analysis. We used the Wilcoxon test (SAS system version 6.07). We counted pollen tubes in each pistil after treatment with the above compounds (20 pistils per treatment) and compared this number with the numbers of pollen tubes in pollinations with tobacco exudate. If the *P* value was ≤0.05, the difference between the two pollination experiments was taken to be significant. Using the following compounds, the average number of penetrated pollen tubes in the stigma and associated *P* values were: tobacco exudate, 21; *Petunia* exudate, 46 (*P* = 0.03); *Lilium* exudate, 0; purified oil, 25 (*P* = 0.86); purified oil:*Lilium* exudate 1:1, 24 (*P* = 0.96); trilinolein, 33 (*P* = 0.28); triolein, 15 (*P* = 0.33); trilinolenin, 11 (*P* = 0.02); saturated triacylglycerides, 1 (*P* = 0.0001); trilinolelaidin, 2 (*P* = 0.0001); glycerol, 0; oleic acid, 8 (*P* = 0.002); linoleic acid, 0; linolenic acid, 0; monolinolein, 0; dilinolein, 3 (*P* = 0.0001); AR20, 0; AR200, 3 (*P* = 0.0001); CR350, 0; DC344, 0; DC200, 1 (*P* = 0.0001).

Transmission electron microscopy (TEM) was performed as described¹⁷.

Scanning electron microscopy (SEM). Fresh pistils were mounted on stubs and sputtered with gold. The samples were examined with a JEOL JSM-T300.

***In vitro* germination.** An artificial boundary between oil and germination medium or between oil and water was created by applying 1 ml germination medium or water (agarose, 0.7%; toluidine blue, 0.01%) to a glass slide. After polymerization, the gel was cut with a razor blade to produce a sharp edge. A coverslip was placed on the gel with an overlap at the edges of 2–3 mm. After this, 50 µl of a suspension of pollen in oil was applied to the edge of the gel. Pollen germination and pollen-tube growth were viewed with a Leitz dialux 20 EB microscope.

Received 26 November 1997; accepted 5 March 1998.

- Goldman, M. H. S., Goldberg, R. B. & Mariani, C. Female sterile tobacco plants are produced by stigma specific cell ablation. *EMBO J.* **13**, 2976–2984 (1994).
- Preuss, D., Lemieux, B., Yen, G. & Davis, R. W. A conditional sterile mutation eliminates surface components from *Arabidopsis* pollen and disrupts cell signalling during fertilization. *Genes Dev.* **7**, 974–985 (1993).
- Dickinson, H. G. Pollen dressed for success. *Nature* **364**, 573–574 (1993).
- Dickinson, H. G. Self-pollination: simply a social disease? *Nature* **367**, 517–518 (1994).
- Taylor, L. P. & Hepler, P. K. Pollen germination and tube growth. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**, 461–491 (1997).
- Mariani, C., De Beuckeleer, M., Truettner, J., Leemans, J. & Goldberg, R. B. Induction of male sterility in plants by a chimeric ribonuclease gene. *Nature* **347**, 737–741 (1990).
- Konar, R. N. & Linskens, H. F. Physiology and biochemistry of the stigmatic fluid of *Petunia hybrida*. *Planta* **71**, 378–387 (1966).
- Labarca, C., Kroh, M. & Loewus, F. The composition of stigmatic exudate from *Lilium longiflorum*. Labelling studies with myo-inositol, D-glucose, and L-proline. *Plant Physiol.* **46**, 150–156 (1970).
- Cresti, M., Keijzer, C. J., Tiezzi, A., Ciampolini, F. & Focardi, S. Stigma of *Nicotiana*: ultrastructural and biochemical studies. *Am. J. Bot.* **73**, 1713–1722 (1986).
- Dumas, C. Lipochemistry of the progamic stage of a self-incompatible species: neutral lipids and fatty acids of the secretory stigma during its glandular activity, and the solid style, the ovary and the anther in *Forsythia intermedia* Zab. (heterostylis species). *Planta* **137**, 177–184 (1977).
- Elleman, C. J., Franklin-Tong, V. & Dickinson, H. G. Pollination in species with dry stigma: the nature of the early stigmatic response and the pathway taken by the pollen tubes. *New Phytol.* **121**, 413–424 (1992).
- Kandasamy, M. K., Nasrallah, J. B. & Nasrallah, M. E. Pollen pistil interactions and developmental regulation of pollen tube growth in *Arabidopsis*. *Development* **120**, 3405–3418 (1994).

13. Evans, D. E., Sang, J. P., Cominos, X., Rothnie, N. E. & Knox, R. B. A study of phospholipids and galactolipids in pollen of two lines of *Brassica napus* L. (rapeseed) with different ratios of linoleic to linolenic acid. *Plant Physiol.* **92**, 418–424 (1990).
14. Piffanelli, P., Ross, J. H. E. & Murphy, D. J. Intra and extracellular lipid composition and associated gene expression patterns during pollen development in *Brassica napus*. *Plant J.* **11**, 549–562 (1997).
15. Dickinson, H. G. Dry stigmas, water and self-incompatibility in *Brassica*. *Sex. Plant Reprod.* **8**, 1–10 (1995).
16. Elleman, C. J. & Dickinson, H. G. Identification of pollen components regulating pollination-specific responses in the stigmatic papillae of *Brassica oleracea*. *New Phytol.* **133**, 197–205 (1996).
17. Wolters-Arts, M., Derksen, J., Kooijman, J. W. & Mariani, C. Stigma development in *Nicotiana tabacum*. Cell death in transgenic plants as a marker to follow cell fate at high resolution. *Sex. Plant Reprod.* **9**, 243–254 (1996).
18. Hülkamp, M., Kopczak, S. D., Horejs, T. F., Kihl, B. K. & Pruitt, R. E. Identification of genes required for pollen-stigma recognition in *Arabidopsis thaliana*. *Plant J.* **815**, 703–714 (1993).
19. Cheung, A. Y., Zhan, X., Wang, H. & Wu, H. Organ-specific and AGAMOUS-related expression and glycosylation of a pollen tube growth-promoting protein. *Proc. Natl Acad. Sci. USA* **93**, 3853–3858 (1996).
20. Lolle, S. J. & Cheung, A. Y. Promiscuous germination and growth of wild type pollen from *Arabidopsis* and related species on the shoot of the *Arabidopsis* mutant, *fiddlehead*. *Dev. Biol.* **155**, 250–258 (1993).
21. Lolle, S. J. *et al.* Developmental regulation of cell interactions in the *Arabidopsis fiddlehead-1* mutant: a role for the epidermal cell wall and cuticle. *Dev. Biol.* **189**, 311–321 (1997).
22. Boden, N. in *Micelles, Membranes, Microemulsions, and Monolayers* (eds Gelbart, W. M., Ben-Shaul, A. & Roux, D.) 153–211 (Springer, New York, 1994).
23. Ikeda, S., Nasrallah, J. B., Dixit, R., Preiss S. & Nasrallah, M. E. An aquaporin like gene required for the *Brassica* self-incompatibility response. *Science* **276**, 1564–1566 (1997).
24. Pruitt, R. E. Molecular mechanics of smart stigmas. *Trends Plant Sci.* **2**, 328–4329 (1997).
25. Baier, H. & Bonhoeffer, F. Axon guidance by gradients of a target-derived component. *Science* **255**, 472–475 (1992).
26. Maret, G. & Dransfeld, K. Biomolecules and polymers in high steady magnetic fields. *Top. Appl. Phys.* **57**, 143–204 (1985).
27. Malho, R., Feijo, J. A. & Pais, M. S. S. Effect of electrical fields and external ionic currents on pollen tube orientation. *Sex. Plant Reprod.* **5**, 57–63 (1992).
28. Reger, B. J., Chaubal, R. & Pressley, R. Chemotropic responses by pearl millet pollen tubes. *Sex. Plant Reprod.* **5**, 47–56 (1992).
29. Robbertse, P. J., Lock, J. J., Stoffberg, E. & Coetzer, L. A. Effect of boron on directionality of pollen tube growth in *Petunia* and *Agapanthus*. *South African J. Bot.* **56**, 487–492 (1990).
30. Hyde, G. J. & Heath, I. B. Ca²⁺-dependent polarization of axis establishment in the tip-growing organism, *Saprolegnia ferax*, by gradients of the ionophore A23187. *Eur. J. Cell Biol.* **67**, 356–362 (1995).

Acknowledgements. We thank A. Pereira for the gift of the *cer* and *pop* mutants, J. Derksen for reading the manuscript and M.B. van Veen for artwork.

Correspondence and requests for materials should be addressed to M.W.-A. (e-mail: mwalters@sci.kun.nl).

Dynamic activation of endothelial nitric oxide synthase by Hsp90

Guillermo García-Cardena^{†*}, Roger Fan^{*}, Vijay Shah[‡], Raffaella Sorrentino[§], Giuseppe Cirino[§], Andreas Papapetropoulos^{*} & William C. Sessa^{*}

^{*} Department of Pharmacology and Molecular Cardiology Program, and [‡] Department of Medicine, Boyer Center for Molecular Medicine, Yale University School of Medicine, New Haven, Connecticut 06536, USA
[§] Dipartimento di Farmacologia Sperimentale, Università degli Studi di Napoli "Federico II", 80131 Napoli, Italy

Heat-shock protein 90 (Hsp90) coordinates the trafficking and regulation of diverse signalling proteins, but its precise role in regulating specific cellular targets is not known^{1,2}. Here we show that Hsp90 associates with endothelial nitric oxide synthase (eNOS) and is rapidly recruited to the eNOS complex by agonists that stimulate production of nitric oxide, namely vascular endothelial growth factor, histamine and fluid shear stress. Moreover, the binding of Hsp90 to eNOS enhances the activation of eNOS. Inhibition of signalling through Hsp90 attenuates both agonist-stimulated production of nitric oxide and endothelium-dependent relaxation of isolated blood vessels. Our results indicate that Hsp90 facilitates signalling mediated by growth-factor, G-protein and mechanotransduction pathways that lead to the activation of eNOS. These observations indicate that in addition

to its role as a molecular chaperone involved in protein folding and maturation, Hsp90 may also be recruited to cellular targets depending on the activation state of the cell.

Several peripheral membrane signalling proteins, including Src and its family members, Raf, G-protein $\beta\gamma$ subunits and mitogen-activated protein kinase kinase (MEK) can exist in heterocomplexes with Hsp90 (ref. 3). However, the regulation and function of the binding of Hsp90 to its native cellular substrates in mammalian cells remains unknown. Genetic studies in *Saccharomyces cerevisiae* and *Drosophila melanogaster* have shown that *hsp90* homologues are essential genes. In *S. cerevisiae*, Hsp82 is an integral component of reconstituted steroid-receptor- and v-Src-dependent signal-transduction events, whereas in *Drosophila* Hsp83 is involved in mediating receptor tyrosine kinase signalling pathways^{4–10}. Potential mechanisms to account for these effects include Hsp90-mediated effects on protein folding, stability and targeting of signalling proteins.

eNOS is the nitric oxide synthase (NOS) isoform that produces endothelium-derived nitric oxide (NO)¹¹. NO is a highly lipophilic gas that is not stored in cells, but is immediately synthesized and released upon stimulus-dependent activation of NOS. NO derived from the endothelium is important in regulating cardiovascular haemodynamics, proliferation of vascular smooth muscle cells, leukocyte adhesion and platelet aggregation¹². eNOS is a peripheral membrane protein in endothelial cells because of its co-translational *N*-myristoylation and post-translational cysteine palmitoylation¹³. Both fatty-acylation reactions are necessary for specific targeting of eNOS to the Golgi, and cysteine palmitoylation is necessary for movement from the Golgi into glycolipid- and cholesterol-rich plasmalemmal microdomains, caveolae^{14,15}. As eNOS coprecipitates with a protein having a relative molecular mass of 90K (refs 16, 17) and as Hsp90 may be involved in several signal-transduction cascades^{7,18}, we studied the possibility that eNOS may exist as a component of an Hsp90 heterocomplex in endothelial cells.

As seen in Fig. 1a, immunoprecipitation of eNOS from bovine microvascular endothelial cell lysates results in the coprecipitation of Hsp90. Immunoprecipitation with non-immune antisera does not result in the precipitation or coprecipitation of the proteins. Immunoprecipitation of Hsp90 also coprecipitates eNOS. As shown for other Hsp90-containing complexes, the eNOS–Hsp90 complex disassembles if immunocomplexes are washed with a high-ionic-strength buffer (0.5 M NaCl). To determine whether eNOS interacts

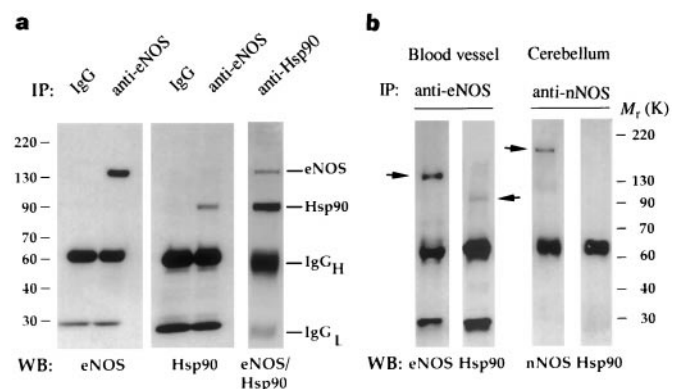


Figure 1 Association between eNOS and Hsp90. **a**, Lysates from BLMVECs were immunoprecipitated with non-immune mouse immunoglobulin (IgG) or with monoclonal antibody against eNOS and Hsp90. Immunoprecipitates (IPs) were western blotted (WB) for eNOS (left), Hsp90 (centre) or both proteins (right). IgG_H and IgG_L denote IgG heavy and light chains, respectively. **b**, Lysates from a single rat aorta or cerebellum were immunoprecipitated with antibodies against eNOS or nNOS. IPs were blotted for eNOS, nNOS and Hsp90. In **b**, the arrows (from left to right) denote eNOS, Hsp90 and nNOS, respectively.

[†] Present address: Harvard Medical School, Vascular Research Division, Brigham and Women's Hospital, 221 Longwood Avenue, LMRC-406, Boston, Massachusetts 02115, USA.

with Hsp90 in intact tissue and whether a structurally distinct, but functionally homologous, NOS isoform associates with Hsp90, we immunoprecipitated eNOS from blood vessels and neuronal NOS (nNOS) from cerebellum. Immunoprecipitation of eNOS from rat aorta coprecipitates Hsp90 (Fig. 1b, left), whereas precipitation of nNOS from cerebellum does not result in the association of these proteins under these conditions (Fig. 1b, right). This demonstrates that eNOS exists in a complex with Hsp90 both in isolated endothelial cells and in the endothelium lining of intact blood vessels.

Next we examined whether the observed cellular interactions between eNOS and Hsp90 were functionally relevant to NOS activation in a heterologous expression system. COS-7 cells were cotransfected with complementary DNAs encoding eNOS and Hsp90 β , and immunocomplex formation and NOS activity were examined. COS-7 cells do not express eNOS but do express Hsp90 (Fig. 2a, left and centre panels). Transfection of COS-7 cells with eNOS cDNA alone results in its expression (Fig. 2a, left panel) and subsequent coprecipitation of the protein with endogenous Hsp90 (Fig. 2a, right panel). Cotransfection of eNOS with Hsp90 increased the total amount of Hsp90 in lysates (centre panel) and the amount of Hsp90 associated with eNOS (right panel). Cotransfection of eNOS with Hsp90 markedly enhances NOS activity in cellular

lysates compared with cells transfected with eNOS alone (Fig. 3b; an increase of $208 \pm 3.8\%$ in NOS activity is seen; $n = 3$; NOS activity is undetectable in mock-transfected COS cells). These data indicate that overexpression of Hsp90 increases the amount of Hsp90 that is bound to eNOS, and that this interaction results in greater NOS activity.

To test directly whether Hsp90 could interact with eNOS, purified proteins were incubated together and immunoprecipitated. As seen in Fig. 3a, Hsp90 can form a complex with eNOS in solution. This interaction results in an increase in eNOS activity (Fig. 3b; an increase of 247 ± 13.5 is seen; $n = 4$, $P < 0.05$). Heat-denatured Hsp90 does not increase NOS activity). Figure 3c shows that Hsp90 activates eNOS in a dose-dependent manner and that this increase in activity is blocked by the specific NOS inhibitor nitro-L-arginine methyl ester (L-NAME; $n = 4$). Thus Hsp90 directly binds to eNOS and facilitates NOS-mediated catalysis *in vitro*.

Nitric oxide is released from endothelial cells in response to growth factors, calcium-mobilizing agonists and shear stress^{19–21}. We therefore studied whether vascular endothelial cell growth factor (VEGF), histamine and fluid shear stress influenced the interaction between eNOS and Hsp90. As shown in Fig. 4a (left and centre panels), VEGF (40 ng ml^{-1}) and histamine ($10 \mu\text{M}$) each stimulate the rapid (with 1 min) recruitment of Hsp90 to eNOS. Exposure of endothelial cells to fluid shear stress (15 dynes cm^{-2}) stimulates the association of eNOS and Hsp90 (Fig. 4a, right panel), but the total amount of cellular Hsp90 does not change (Fig. 4a, bottom right panel). When VEGF and histamine were used, the enhanced interaction between eNOS and Hsp90 corresponded with the functional measurement of NO production at 1 min and 5 min after stimulation (control, 47.0 ± 9.1 and $49.7 \pm 3.3 \text{ pmol cGMP mg}^{-1}$ protein; VEGF, $94.8 \pm 16.1^*$ and $94.5 \pm 7.5^*$; and histamine, 136.7 ± 47.4 and $155 \pm 0.9^*$ at 1 and 5 min, respectively; $n = 3$; asterisks indicate that these values are significantly different from control values; $P < 0.05$). The rapid stimulus-dependent formation of the eNOS–Hsp90 heterocomplex indicates that this interaction precedes or occurs simultaneously with other signalling events, such as calcium mobilization or activation of other downstream effectors required for NO release.

To test whether a specific stimulus could facilitate the ‘activation’ of Hsp90 and the productive interaction of Hsp90 with eNOS, we investigated whether lysates prepared from control or VEGF-activated ECV 304 cells (an immortalized human endothelial cell line²²) could transactivate purified eNOS. In preliminary studies, NOS activity or eNOS protein could not be detected in lysates from ECV 304 cells ($n = 3$ in direct NOS activity and cyclic GMP measurements, and $n = 4$ for western blotting, with lysates from bovine lung microvascular cells serving as a positive control), although these cells

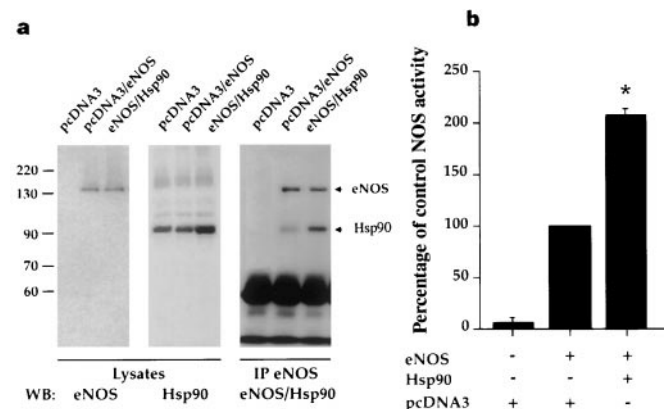


Figure 2 Overexpression of Hsp90 increases eNOS activity. **a**, COS cells were transfected with vector alone (pcDNA3) or with vector plus eNOS and Hsp90 cDNAs and the expression and association of both proteins was determined. **b**, NOS activity was measured in lysates from the same experiments as in **a**. Data are mean $\pm$ s.e.m., $n = 3$ experiments, done in duplicate; asterisk denotes $P < 0.05$.

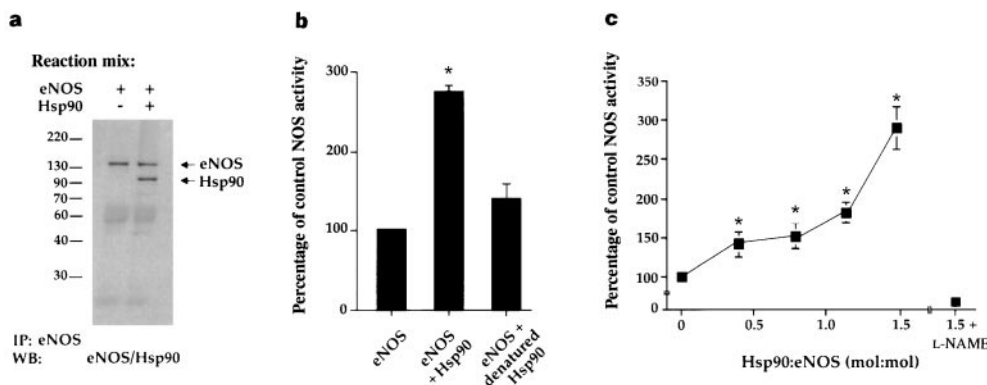


Figure 3 Direct interaction and activation of eNOS by Hsp90. **a**, Purified eNOS (7.4 pmol) was incubated without or with purified Hsp90 (11 pmol) and the interaction between them determined. **b**, eNOS (7.4 pmol) was incubated in the absence or presence of native or heat-denatured Hsp90 (11 pmol) and NOS activity was determined. **c**, Increasing amounts of Hsp90 were incubated with a fixed amount of eNOS and NOS activity was determined. At the highest ratio of Hsp90:eNOS (1.42), the proteins were incubated with L-NAME (1 mM). Data are mean $\pm$ s.e.m., $n = 4$ experiments in duplicate; asterisk denotes $P < 0.05$.

express ample amounts of Hsp90. Incubation of VEGF-activated lysates, but not control lysates, with purified eNOS results in an increase in NOS activity (Fig. 4b). The increase in NOS activity is markedly attenuated compared with control levels by immunodepletion of Hsp90 (using a specific anti-Hsp90 monoclonal antibody, but not with a non-immune IgG; $n = 3$ experiments in duplicate) from VEGF-activated ECV 304 lysates. These data indicate that VEGF signalling may activate Hsp90 or an Hsp90-associated protein, thereby increasing its affinity for eNOS, and that this association stimulates NOS activation.

To study the relationship between Hsp90-mediated signalling and NO production, we used specific inhibitor of Hsp90-mediated processes, the ansamycin antibiotic geldanamycin. Geldanamycin binds to the ATP-binding site of Hsp90 and influences the conformational stability of Hsp90 binding to its substrates^{23,24,30}. Treatment of cultured endothelial cells with geldanamycin ($1 \mu\text{g ml}^{-1}$) attenuates histamine- and VEGF-stimulated cGMP production (Fig. 5a), but does not influence the ability of the NO donor sodium nitroprusside ($100 \mu\text{M}$) to increase cGMP production (21.5 ± 6.9 and 18.7 ± 5.3 pmol cGMP per mg protein produced in the absence and presence of geldanamycin, respectively; $n = 6$). In addition, geldanamycin ($1 \mu\text{g ml}^{-1}$) does not directly inhibit NOS activity of endothelial cell lysates, whereas L-NAME does (9.1 ± 0.4 , 8.1 ± 0.17 and 0.19 ± 0.63 nmol per mg per min of eNOS activity for control, geldanamycin-treated and L-NAME-treated lysates, respectively; $n = 3$), indicating that inhibition of

agonist-induced NO release was likely to have been due to inhibition of Hsp90-mediated signalling.

To test the involvement of Hsp90 in NO-dependent relaxation of intact blood vessels, we examined the influence of geldanamycin on acetylcholine-induced relaxations of rat aorta. Mechanistically, both histamine and acetylcholine stimulate NO synthesis in a calcium-dependent manner, and acetylcholine-induced vasodilation is NO-mediated as NOS inhibitors abrogate endothelium-dependent relaxations in rat aortic rings^{25,26}. As shown in Fig. 5b, geldanamycin ($10 \mu\text{g ml}^{-1}$) blocks acetylcholine-induced vasorelaxation of isolated rat aortic rings. The effects of geldanamycin are reversible, as washout of geldanamycin restores acetylcholine-induced vasorelaxation (Fig. 5b: far right in bioassay trace). Moreover, geldanamycin significantly attenuates the maximal response to acetylcholine (with effects similar in magnitude to those of NOS inhibitors²⁵) and does not influence the direct relaxant effects of the NO-donor nitroglycerine (Fig. 5c). The data from Figs 4 and 5 show

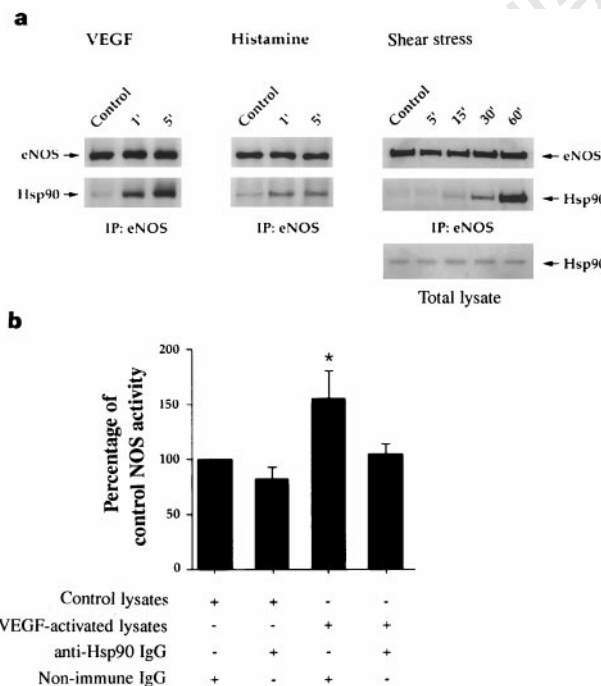


Figure 4 The stimulus-dependent association of Hsp90 and eNOS. **a**, HUVECs were treated with VEGF or histamine (left and centre panels) and eNOS was immunoprecipitated (IP). Western blot (WB) analysis of the presence of eNOS and Hsp90 was performed. In the right panel; BLMVEC were exposed to shear stress for the indicated times and processed as described. The lower blot shows the levels of Hsp90 in total cell lysates. **b**, Control or VEGF-treated ECV 304 cell lysates were incubated with non-immune IgG or the anti-Hsp90 monoclonal antibody and samples were immunoprecipitated. Antibody-treated samples were transferred to a separate tube containing NOS, and NOS activity was determined. Data are presented as mean $\pm$ s.e.m., $n = 3$; asterisk represents significant differences ($P < 0.05$).

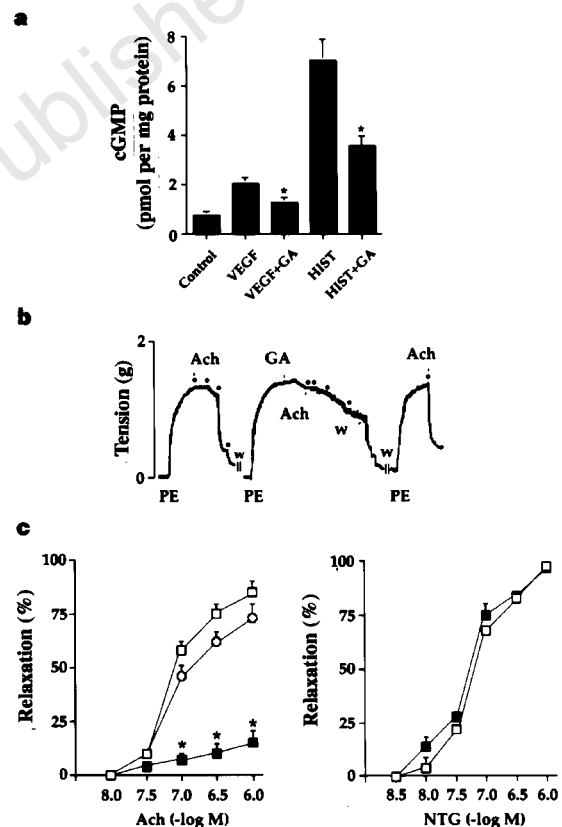


Figure 5 Role of Hsp90 in stimulated NO release and endothelium-dependent relaxations of rat aorta. **a**, HUVECs were pretreated with geldanamycin (GA) or vehicle and stimulated with VEGF or histamine (HIST). Levels of cGMP were then determined. The data are mean $\pm$ s.e.m., $n = 8-12$; asterisks represent significant differences ($P < 0.05$). **b**, Bioassay trace of acetylcholine (Ach)-induced endothelium-dependent relaxations. The aortic ring was precontracted with phenylephrine (PE) and a cumulative dose-response curve to Ach was examined (Ach concentrations were 30 nM, 100 nM, 300 nM and $1 \mu\text{M}$; first four solid circles). Drugs were then washed out (w) for 15 min and the vessel was precontracted with PE again. GA ($10 \mu\text{g ml}^{-1}$) was added to the organ bath for 15 min and the Ach dose-response repeated (second four solid circles). The rings were washed again (w) for 15 min, precontracted, and a single dose of Ach (300 nM) was added to demonstrate the reversibility of the GA blockage. A summary of the bioassay experiments is shown in **c**. GA significantly attenuates Ach-induced vasorelaxation (left panel) but not nitroglycerin-induced vasorelaxation (NTG, right panel; means $\pm$ s.e.m., $n = 4$; asterisks denote $P < 0.05$). Open boxes, control; filled boxes, in the presence of GA; and open circles (left panel) after GA washout.

that Hsp90 associates with eNOS in a stimulus-dependent manner and indicate that Hsp90 may be necessary for NO release and endothelium-dependent responses in intact blood vessels.

Our data show that Hsp90 is a physiological binding partner and regulator of eNOS. We predict that, upon activation of the endothelium by VEGF, histamine or fluid shear stress, the 'molecular attraction' of eNOS and/or Hsp90 for the other protein increases, resulting in the association of the two proteins and subsequent activation of eNOS. On the basis of our *in vitro* results, we suggest that Hsp90 may act as an allosteric modulator of eNOS by inducing a conformational change in the enzyme or by stabilizing the dimeric form. In the cellular context, direct modulation of eNOS may be accompanied by Hsp90 serving as a scaffolding protein for organization of other associated regulatory proteins that are recruited to the NOS complex. Thus, Hsp90 can be considered to be a new intermediate in the growth-factor, agonist and mechanical signalling cascades leading to NO release. □

Methods

Immunoprecipitation and immunoblot analysis. Cells, blood vessels and cerebella were lysed on ice in the following lysis buffer: 20 mM Tris-HCl, pH 7.4, 2.5 mM EDTA, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS, 100 mM NaCl, 10 mM NaF, 1 mM Na_3VO_4 , 1 mM Pefabloc, $10 \mu\text{g ml}^{-1}$ aprotinin, $10 \mu\text{g ml}^{-1}$ leupeptin, $10 \mu\text{g ml}^{-1}$ pepstatin A. Protein concentrations were determined using a Lowry assay. SDS-PAGE and immunoblotting were performed using standard procedures. Monoclonal antibodies used in immunoprecipitations and western blotting include: an anti-eNOS antibody (H32); an anti-nNOS antibody; and anti-HSP90 antibodies (StressGen SPA-845, SPA-830). Proteins were detected by enhanced chemiluminescence.

Cell transfections. The bovine eNOS and human Hsp90 β cDNAs (from ATCC) were cloned into the mammalian expression vector pcDNA3 as described¹⁵. COS-7 cells were plated and transfected with vector alone, eNOS or Hsp90 using lipofectamine (GIBCO) according to the manufacturer's protocol. On day 3 after transfection, the expression of appropriate proteins was confirmed.

NOS-activity assays. The conversion of ^3H -labelled L-arginine to ^3H -labelled L-citrulline was used to determine NOS activity in COS cell lysates or with *Escherichia coli*-derived purified eNOS as described¹⁶. The specific activities of different batches of eNOS were 70–200 nmol per mg per min. In experiments examining the effects of Hsp90 on eNOS activity, eNOS was preincubated with Hsp90 α (StressGen Lab.) for 15 min at room temperature, and the reaction was initiated by the addition of NOS cofactors and incubated for an additional 10 min at 37°C.

Transactivation experiments. ECV 304 cells were treated with VEGF or control buffer for 30 s and lysates were prepared in lysis buffer. Equal amounts of protein were incubated with mouse non-immune IgG or with anti-Hsp90 monoclonal antibody for 3 h at 4°C. Immune complexes were removed with protein-A-Sepharose beads and equal amounts of protein were transferred to a tube containing purified eNOS. NOS activity was assessed as described.

Cell activation and cGMP production. Human umbilical vein endothelial cells (HUVECs) and bovine lung microvascular endothelial cells (BLMVECs) were cultured as described^{16,27}. HUVECs were grown to confluence in culture dishes (for experiments with VEGF and histamine) and BLMVECs were grown on glass coverslips (shear stress experiments). The growth medium was replaced 18 h before the experiment with serum-free, growth-factor-free medium supplemented with 2% bovine serum albumin. Human VEGF₁₆₅ (40 ng ml^{-1}) or histamine ($10 \mu\text{M}$) was added at different time points. Laminar fluid shear stress (15 dynes cm^{-2}) was generated using a parallel plate chamber through which serum-free medium was continually perfused through a flow loop as described²⁸. Stationary cultures of slides were kept in dishes containing a controlled volume of serum-free medium.

For cGMP measurements at early time points, HUVECs were cultured on glass coverslips and incubated with rat aortic vascular smooth muscle cells in the absence or presence of the agonist for the indicated times. Levels of cGMP were determined by radioimmune assay (RIA). In experiments examining the effects of geldanamycin on cGMP accumulation, HUVECs were preincubated with geldanamycin or vehicle (dimethyl sulphoxide, 0.1%) for 30 min; this was

followed by stimulation with VEGF and histamine for 5 min, and cGMP accumulation in HUVECs was measured by RIA.

Vasorelaxation of aortic rings. Isometric tension was measured in rat descending thoracic aortic rings as described²⁹. After equilibration of the vessels, the effector concentration for 80% response (EC_{80}) dose of phenylephrine ($0.3 \mu\text{M}$) was added to precontract the rings, and acetylcholine was added in a cumulative manner (from 10 nM to $1 \mu\text{M}$). Endothelium-dependent relaxations were observed. Geldanamycin was added for 15 min after a stable phenylephrine-induced contraction was obtained and acetylcholine-induced relaxations were repeated. An identical protocol was used to assess the direct relaxant effects of the NO donor nitroglycerine (1 nM to $1 \mu\text{M}$).

Statistics. Data are expressed as mean $\pm$ s.e.m. Comparisons were made using a two-tailed student's *t*-test or analysis of variance (ANOVA). Differences were considered to be significant at $P < 0.05$.

Received 24 December 1997; accepted 9 February 1998.

- Rutherford, S. L. & Zucker, C. S. Protein folding and the regulation of signaling pathways. *Cell* **79**, 1129–1132 (1994).
- Johnson, J. L. & Craig, E. A. Protein folding *in vivo*: unraveling complex pathways. *Cell* **90**, 201–204 (1997).
- Pratt, W. B. The role of the Hsp90-based chaperone system in signal transduction by nuclear receptors and receptors signaling via MAP kinase. *Annu. Rev. Pharmacol. Toxicol.* **37**, 297–326 (1997).
- Pratt, W. B. & Toft, D. O. Steroid receptor interactions with heat shock protein and immunophilin chaperones. *Endocrinol. Rev.* **18**, 306–360 (1997).
- Bohen, S. P., Krall, A. & Yamamoto, K. R. Hold'em and fold'm: chaperones and signal transduction. *Science* **268**, 1303–1304 (1995).
- Nathan, D. F. & Lindquist, S. Mutational analysis of Hsp90 function: interactions with a steroid receptor and a protein kinase. *Mol. Cell. Biol.* **15**, 3917–3925 (1995).
- Xu, Y. & Lindquist, S. Heat-shock protein Hsp90 governs the activity of pp60^{src} kinase. *Proc. Natl Acad. Sci. USA* **90**, 7074–7078 (1993).
- Cutforth, T. & Rubin, G. M. Mutations in Hsp83 and cdc37 impair signaling by the Sevenless receptor tyrosine kinase in *Drosophila*. *Cell* **77**, 1027–1036 (1994).
- Doyle, H. J. & Bishop, J. M. Torso, a receptor tyrosine kinase required for embryonic pattern formation, shares substrates with the Sevenless and EGF-R pathways in *Drosophila*. *Genes Dev.* **7**, 633–646 (1993).
- Van der Straten, A., Rommel, C., Dickson, B. & Hafen, E. The heat shock protein 83 (Hsp83) is required for Raf-mediated signaling in *Drosophila*. *EMBO J.* **16**, 1961–1969 (1997).
- Nathan, C. & Xie, Q.-W. Nitric oxide synthases: roles, tolls, and controls. *Cell* **78**, 915–918 (1994).
- Moncada, S., Palmer, R. M. J. & Higgs, E. A. Nitric oxide: physiology, pathophysiology and pharmacology. *Pharmacol. Rev.* **43**, 109–142 (1991).
- Liu, J., Garcia-Cardeña, G. & Sessa, W. C. Biosynthesis and palmitoylation of endothelial nitric oxide synthase: mutagenesis of palmitoylation sites, cysteines-15 and/or -26, argues against depalmitoylation-induced translocation of the enzyme. *Biochemistry* **34**, 12333–12340 (1995).
- Garcia-Cardeña, G., Oh, P., Liu, J., Schnitzer, J. E. & Sessa, W. C. Targeting of nitric oxide synthase to endothelial cell caveolae via palmitoylation: implications for nitric oxide signaling. *Proc. Natl Acad. Sci. USA* **93**, 6448–6453 (1996).
- Liu, J., Hughes, T. E. & Sessa, W. C. The first 35 amino acids and fatty acylation sites determine the molecular targeting of endothelial nitric oxide synthase into the Golgi region of cells: a green fluorescent protein study. *J. Cell Biol.* **137**, 1525–1535 (1997).
- Garcia-Cardeña, G., Fan, R., Stern, D. F., Liu, J. & Sessa, W. C. Endothelial nitric oxide is regulated by tyrosine phosphorylation and interacts with caveolin-1. *J. Biol. Chem.* **271**, 27237–27240 (1996).
- Venema, V. J., Marrero, M. B. & Venema, R. C. Bradykinin-stimulated protein tyrosine phosphorylation promotes endothelial nitric oxide synthase translocation to the cytoskeleton. *Biochem. Biophys. Res. Commun.* **155**, 801–807 (1996).
- Courtneidge, S. A. & Bishop, J. M. Transit of pp60^{src} to the plasma membrane. *Proc. Natl Acad. Sci. USA* **79**, 7117–7121 (1982).
- Sessa, W. C. The nitric oxide synthase family of proteins. *J. Vasc. Res.* **31**, 131–143 (1994).
- Morbidelli, L. *et al.* Nitric oxide mediates mitogenic effect of VEGF on coronary venular endothelium. *Am. J. Physiol.* **270**, H411–H415 (1996).
- Davies, P. F. Flow-mediated endothelial mechanotransduction. *Physiol. Rev.* **75**, 519–560 (1995).
- Takahashi, K., Sawasaki, Y., Hata, J.-I., Mukai, K. & Goto, T. Spontaneous transformation and immortalization of human endothelial cells *in vitro*. *Cell. Dev. Biol.* **25**, 265–274 (1990).
- Stebbins, C. E. *et al.* Crystal structure of an Hsp90-geldanamycin complex: targeting of a protein chaperone by an antitumor agent. *Cell* **89**, 239–250 (1997).
- Grenert, J. P. *et al.* The amino-terminal domain of heat shock protein 90 (hsp90) that binds geldanamycin is an ATP/ADP switch domain that regulates hsp90 conformation. *J. Biol. Chem.* **272**, 23843–23850 (1997).
- Rees, D. D., Palmer, R. M. J., Schulz, R., Hodson, H. F. & Moncada, S. Characterization of three inhibitors of endothelial nitric oxide synthase *in vitro* and *in vivo*. *Br. J. Pharmacol.* **101**, 746–752 (1990).
- Nagao, T., Illiano, S. & Vanhoutte, P. M. Heterogeneous distribution of endothelium-dependent relaxations resistant to N^G -nitro-L-arginine in Rats. *Am. J. Physiol.* **263**, H1090–H1094 (1992).
- Gimbrone, M. A. J. in *Progress in Hemostasis and Thrombosis* (ed. Spaet, T. H.) 1–28 (Grune and Stratton, New York, 1976).
- Frangos, J. A., Eskin, S. G., McIntire, L. V. & Ives, C. L. Flow effects on prostacyclin production by cultured human endothelial cells. *Science* **227**, 1477–1479 (1985).
- Furchtgott, R. F. & Zawadzki, J. V. The obligatory role of endothelial cells in the relaxation of arterial smooth muscle by acetylcholine. *Nature* **288**, 373–376 (1980).
- Prodromou, C. *et al.* Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. *Cell* **90**, 65–75 (1997).

Acknowledgements. We thank J. Pollock for anti-eNOS monoclonal antibody; B. Mayer for nNOS polyclonal antibody; P. Martasek and B. S. Masters for recombinant eNOS; and R. Renteria and P. Kodaman for assistance in the nNOS study. This work was supported by grants from the NIH and the American Heart Association. W.C.S. is an Established Investigator of the American Heart Association.

Correspondence and requests for materials should be addressed to W.C.S. (e-mail: william.sessa@yale.edu).

Defective meiosis in telomere-silencing mutants of *Schizosaccharomyces pombe*

Elaine R. Nimmo^{*†}, Alison L. Pidoux[†], Paul E. Perry[†] & Robin C. Allshire[†]

^{*} Cancer Research Campaign Project, [†] MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

During meiotic prophase, chromosomes frequently adopt a bouquet-like arrangement, with their telomeres clustered close to the nuclear periphery¹⁻³. A dramatic example of this occurs in the fission yeast, *Schizosaccharomyces pombe*, where all telomeres aggregate adjacent to the spindle pole body (SPB)⁴⁻⁷. Nuclei then undergo rapid traverses of the cell, known as 'horsetail' movement, which is led by the SPB dragging telomeres and chromosomes behind^{4,6,7}. This process may initiate or facilitate chromosome pairing before recombination and meiosis. With the aim of identifying components involved in telomere structure and function, we report here the isolation of *S. pombe* mutants defective in the ability to impose transcriptional silencing on genes placed near telomeres⁸. Two of these mutants, *lot2-s17* and *lot3-uv3*, also display a dramatic lengthening of telomeric repeats. *lot3-uv3* carries a mutation in Taz1 (ref. 9), a telomere-binding protein containing a Myb-like motif similar to two human telomere-binding proteins^{10,11}. Meiosis is aberrant in these mutant yeast strains, and our analysis demonstrates a decreased association of telomeres with the SPB in meiotic prophase. This results in defective 'horsetail' movement, a significant reduction in recombination, low spore viability and chromosome missegregation through meiosis.

In fission yeast, functional marker genes are rendered transcriptionally silent by a process reminiscent of position-effect variegation in *Drosophila melanogaster*¹², when they are placed within centromeric DNA^{13,14}, next to the silent mating type loci¹⁵, or adjacent to telomeres⁸. A multiply marked fission yeast strain was constructed

that allowed the selection of mutants that affect telomere-imposed silencing but not silencing within a centromere or the silent mating type locus (Fig. 1a, b).

Mutants displaying defective telomeric silencing were assigned to three complementation groups and were examined for alterations in telomere repeat-array length. Two mutants, designated *lot2-s17* and *lot3-uv3* (for lengthening of telomeres), show dramatic increases (5–10 fold) in telomere length and in intensity of signal obtained with a probe detecting telomere repeats (Fig. 1c) like *taz1Δ*⁹. The third mutant, designated *rat1-s5* (for repression alleviated at telomeres), has no apparent defect in telomere-length regulation because a wild-type terminal fragment of 0.6–0.8 kilobases (kb) (composed of approximately 0.3 kb of terminal repeats and 0.48 kb of telomeric *his3*⁺) is generated (Fig. 1c). Further analysis revealed that *lot3-uv3* is an allele of *taz1*⁺ (ref. 9) and carries a missense mutation (alanine replaced by valine at position 606; A606V) in the putative DNA-recognition helix (helix 3) of the Myb-like telomere DNA-binding motif of Taz1 protein⁹. The products of the *lot2*⁺ and *rat1*⁺ genes remain to be characterized, but neither is an allele of *taz1*⁺.

In addition to defects in silencing and telomere-length regulation, the strains bearing the *lot2-s17* and *lot3-uv3* (henceforth known as *taz1-uv3*) mutations display defective meiosis upon microscopic examination of mating cells. A high incidence of asci with abnormal spore number and/or size was observed (Fig. 2a, b), like the *S. pombe* *rec8* mutant that has defective meiotic sister-chromatid cohesion and recombination¹⁶. The viability of *lot2-s17* and *taz1-uv3* spores was also reduced relative to wild-type and *rat1-s5* (Fig. 2b). Fluorescence-activated cell sorting (FACS) analysis of nuclear content and monitoring of mini-chromosome segregation to *lot2-s17* spores indicated that there were defects in meiotic chromosome segregation (data not shown). The threefold increase in missegregation observed in *lot2-s17* meiosis is an underestimate, because loss can only be assayed in viable spores and aneuploidy (apart from disomic chromosome III) is not tolerated¹⁷. Thus, the reduced viability of these spores is likely to be due to chromosome missegregation during meiosis.

As telomeres might be important in the reorganization of the fission yeast nucleus in preparation for meiosis⁴⁻⁷, this nuclear remodelling may be disrupted in the *lot2-s17* and *taz1-uv3* mutants.

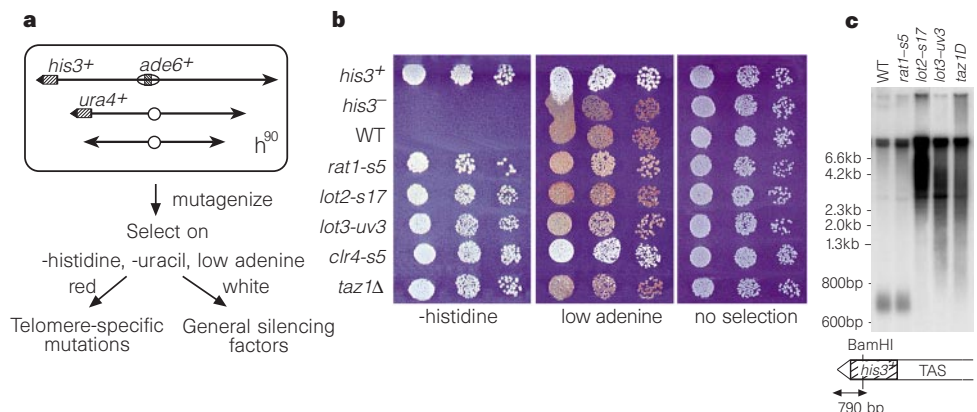


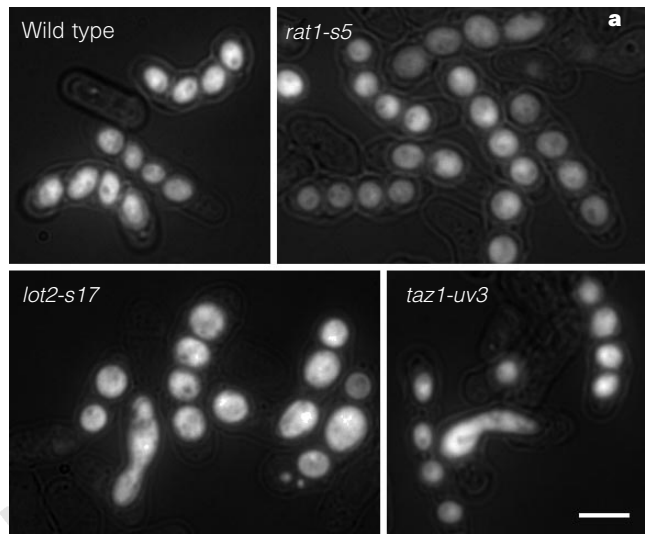
Figure 1 Identification and analysis of *S. pombe* mutants with defective silencing at telomeres. **a**, Screen for mutations allowing the expression of telomeric markers. FY1862 (wild type in all figures), carries *his3*⁺ and *ura4*⁺ approximately 300 bp from the left telomeres of chromosomes I and II respectively, and *ade6*⁺ within *cen1*. Although these genes are functional, the strain is unable to grow on plates lacking uracil or histidine and forms red colonies on low adenine because of transcriptional repression of genes adjacent to telomeres or within centromeres^{8,13,14}. Mutations that derepress telomeric markers should allow growth on plates lacking uracil or histidine. Spontaneous and UV-generated

mutants were isolated and tested for expression of the centromeric *ade6*⁺ gene (red colonies, repressed; white, expressed) or the silent mating-type loci (iodine staining^{15,30}) to assign them to telomere-specific or general silencing factor classes. **b**, Serial dilution assay on no histidine, low adenine or non-selective plates. **c**, Genomic DNA of telomere-specific mutants was digested with *Bam*HI, which in wild-type cuts about 0.79 kb from the terminus within the telomeric *his3*⁺ gene (484 bp of *his3*⁺ and about 300 bp of telomere repeats) and analysed by Southern blotting. The resulting filter was probed with α -³²P-labelled telomere repeat DNA. For comparison, DNA from *taz1Δ* strain was also analysed⁹.

The position of telomeres in meiotic prophase of wild-type, *rat1-s5*, *lot2-s17* and *taz1-uv3* cells was examined by fluorescence *in situ* hybridization^{4,18–20} (FISH) relative to the SPB, detected with antibodies against Sad1 (ref. 21) (Fig. 3a, b). In more than 85% of wild-type and *rat1-s5* cells, all telomeres detected (red signal in Fig. 3) were found localized next to the SPB (green signal). In the *lot2-s17* and *taz1-uv3* mutants, however, telomeres were associated with the SPB in only 4.7 and 39% of cells, respectively (Fig. 3b). Because only a small number of telomere signals were detected in *lot2-s17* and *taz1-uv3*, telomeres may not be completely unclustered. However, it is clear that these mutants are defective either in the establishment or in the maintenance of telomere aggregation at the SPB.

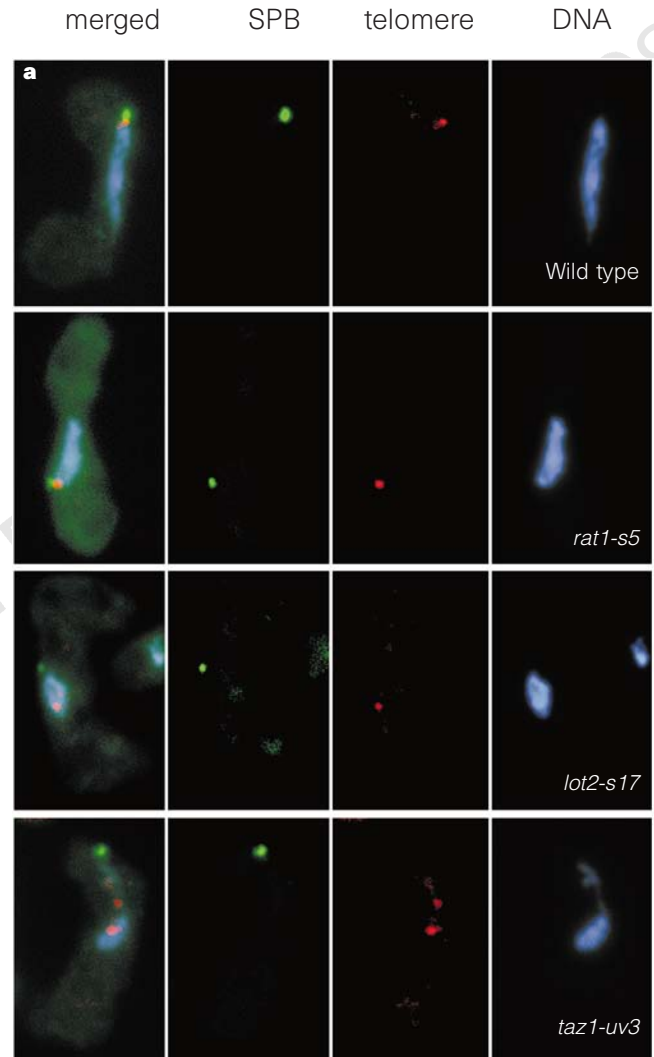
If the association of telomeres with the SPB is important for meiotic prophase horsetail movement, then the *lot2-s17* and *taz1-uv3* cells should display aberrant nuclear dynamics at this stage. Nuclear horsetail movement was analysed in live wild-type and mutant zygotes expressing a functional version of the Swi6 protein fused to green fluorescent protein²² (GFP; Fig. 4a, b). Swi6 is localized at fission yeast centromeres, telomeres and the silent mating type loci¹⁹; GFP–Swi6 shows a localization pattern identical to endogenous Swi6 (A.L.P. and R.C.A., unpublished results). During meiotic prophase in wild-type zygotes, the characteristic horsetail-shaped nucleus can be seen owing to the low level of GFP–Swi6 in the nucleoplasm; the whole nucleus moves around the cell,

as has been reported previously for Hoechst 33342-stained cells^{4,7}, a process that depends on intact microtubules²³. In addition, there is a bright spot of fluorescence at the leading tip of movement, which presumably represents telomeres clustered at the SPB. Smaller spots, likely to be centromeres, are seen trailing in the body of the nucleus



	Asci with 4 normal spores	Asci with abnormal number / size spores					Spore viability (%)
		1	2	3	4	>4	
Wild type	98	0	1	1	0	0	53.6
<i>rat1-s5</i>	95	1	3	0	0	1	61.1
<i>lot2-s17</i>	26	38	22	3	10	1	7.4
<i>taz1-uv3</i>	47	10	23	6	14	0	24.8

Figure 2 Telomere mutants display defects in meiosis. **a**, Spores from wild-type, *rat1-s5*, *lot2-s17* and *taz1-uv3* cells stained with propidium iodide. Wild-type and *rat1-s5* have normal four-spored asci, whereas many aberrant asci with abnormal spore numbers or size are seen in *lot2-s17* and *lot3/taz1-uv3*. **b**, Frequency of asci having 1, 2, 3, 4 and more than 4 spores (100 of each were scored) and spore viability (spore viability was 50–60% in all wild-type crosses tested).



	Anti-Sad1 SBP Telomere FISH				
Wild type	85.5%	6.8%	2.6%	5.1%	n=117
<i>rat1-s5</i>	95.4%	1.5%	1.5%	1.5%	n=131
<i>lot2-s17</i>	4.7%	56.5%	37.6%	1.2%	n=85
<i>taz1-uv3</i>	39.3%	43.8%	6.2%	10.7%	n=112

Figure 3 Telomeres fail to associate with the spindle pole body in telomere mutants. **a**, Fluorescence immunolocalization and FISH were performed on meiotic prophase cells to detect the relative positions of the spindle pole body (SPB) and the telomeres. The merged image is shown together with the Sad1 (SPB; green), telomere (red) and DNA (blue) panels. Width of panels, 60 μ m. **b**, Cells were categorized on the basis of relative positions of anti-Sad1 and telomere FISH signals; percentages are shown for the wild-type, *rat1-s5*, *lot2-s17* and *taz1-uv3* cells.

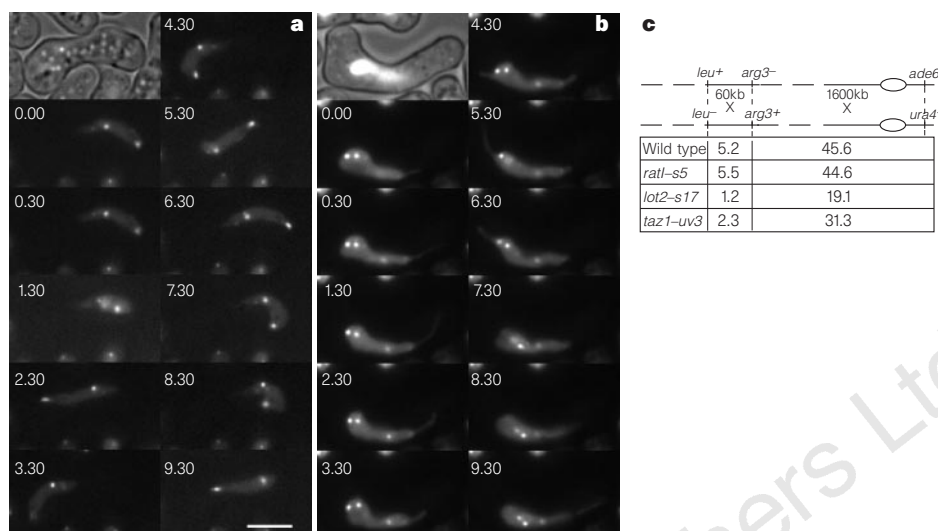


Figure 4 *lot2-s17* displays aberrant horsetail movement and reduced recombination during meiotic prophase. Live analysis of nuclear movement in **a**, wild-type, and **b**, *lot2-s17* zygotes expressing GFP-Swi6. The first image of each set is combined brightfield/fluorescence. Fluorescence images were obtained at the times indicated in minutes. In *lot2-s17*, a thin strand of nucleoplasm is seen extending from the body of the nucleus. There is no bright spot of GFP-Swi6 at the leading tip, however, and the bulk of the nucleus remains relatively stationary. Scale bar, 5 μ m. (See also Supplementary Information.) **c**, Recombination is

reduced in *lot2-s17* and *taz1-uv3*. The frequency of recombination across two intervals on the left arm of chromosome I was calculated from homozygous crosses of wild-type, *rat1-s5*, *lot2-s17* and *taz1-uv3* cells of opposite mating type and bearing either the marker combination shown on the top (*leu*⁺, *arg*³-, *ade*6⁺ / *ura*4⁻) or the lower lines (*leu*⁻, *arg*3⁺, *ade*6⁻ / *ura*4⁺). The recombination frequencies in the 60-kb *leu*-*arg*3 (percentage *leu*⁺ *arg*⁺ progeny) and 1,600-kb *arg*3-*ade*6 (percentage *arg*⁺ *ade*⁺ progeny) intervals is shown in the second and third columns, respectively.

(Fig. 4a). In contrast, in *lot2-s17* zygotes only a thin extrusion of nucleoplasm is observed moving backwards and forwards, while the bulk of the nucleus remains relatively stationary (Fig. 4b). No bright fluorescent spot is observed at the leading tip, although a small weak spot is sometimes seen. Together, live analyses (see Supplementary Information) and our FISH/SPB telomere-localization data indicate that *lot2-s17* is defective in telomere aggregation at the SPB, and that although the nucleus undergoes horsetail-like movements, this fails to involve the chromosomes because they are not attached by their telomeres to the SPB. Similar defects in nuclear dynamics were seen during the horsetail stage in a smaller proportion of *taz1-uv3* cells (see Supplementary Information), consistent with the more frequent telomere-SPB associations observed in this strain (Fig. 3b).

If the association of telomeres with the SPB in the horsetail phase of meiotic prophase is required to facilitate pairing of homologous chromosomes before the first meiotic division, a reduction in pairing should occur in *lot2-s17* and *taz1-uv3* cells, which should manifest as a lower rate of recombination between homologous chromosomes. The frequency of recombination was measured across a 60-kb interval on the left arm of chromosome I and an adjacent interval covering 1,600 kb to the centromere (Fig. 4c). In *rat1-s5*, the genetic distance in both these intervals is approximately the same as in wild type. However, the frequency of recombination in the 60-kb region is reduced in *lot2-s17* and *taz1-uv3* by 4.3- and 2.3-fold, respectively. On average, 19 recombination events occur on chromosome I per meiosis²⁴. As chromosome I occupies 5,600 kb^{25,26} and, assuming a random distribution, five recombination events per meiosis should occur in the 1,600-kb interval. The reduction observed in this interval in *lot2-s17* indicates that a recombination event may now occur in only 19% of meioses. If this is a single exchange event, the true recombination frequency in the 1,600-kb interval in the viable spores may therefore be reduced up to 24-fold in *lot2-s17*, and 14-fold in *taz1-uv3*.

The reorganization of the fission yeast nucleus at meiotic prophase so that telomeres aggregate at the SPB, and chromosomes are pulled around by the dynamic behaviour of the SPB, appears to be

required to facilitate chromosome pairing, recombination and meiosis. Defective SPB structure and function also leads to similar defects in meiosis²⁷. The mechanism by which telomeres move to and are maintained at the SPB is independent of SPB/nuclear movement however, which still occurs (at a similar rate) in *lot2-s17* and *taz1-uv3*. The *lot2-s17* and *taz1-uv3* mutants are apparently defective in moving, gathering or holding telomeres to the SPB. Components involved in telomere structure and length regulation are therefore required for these events to take place.

Nuclei of explanted rat spermatocytes undergo a period of oscillatory chromosome movement²⁸ which correlates with the late leptotene-early zygotene stages at which telomeres in maize, rodents and humans cluster to one region of the nuclear envelope¹⁻³. Telomere components such as TRF1 or TRF2 in mammals^{10,11} may be required for this meiotic nuclear reorganization. In humans, aberrant recombination is frequently associated with incidences of aneuploidy such as trisomy 21 (ref. 29). Defects in human telomere structure and function may be responsible for altered recombination frequency or distribution, leading to chromosome missegregation, aberrant meiosis and reduced fertility. □

Methods

Strain construction and manipulation. *his*3⁺ and *ura*4⁺ genes were placed adjacent to telomeres by transformation of FY1718 (*h*⁻, *his*3-D1, *leu*1-32, *ura*4/4DS/E, *ade*6-210, *clr*4-s5) with appropriate TAS-*his*3⁺/*ura*4⁺-TEL DNA fragments. The *clr*4 mutation allowed some *his*3⁺/*ura*4⁺ expression at telomeres. *His*⁺ *ura*⁺ prototrophs were crossed to a wild-type strain with *ade*6⁺ in *cen*1 and FY1862 isolated (*h*⁹⁰, *ade*6-D1, *his*3-D1, *leu*1-3, *ura*4-D18, *otr*1R(*Sph*I)::*ade*6⁺ TAS-*his*3⁺-*tel*1(L), TAS-*ura*4⁺-*tel*2(L)). Genotypes were confirmed by southern blotting, pulsed-field gel electrophoresis and test crosses. The *Saccharomyces cerevisiae* *LEU*2 gene (complements *leu*1-32) was amplified by PCR with 76-bp and 77-bp tails of homology to a region on chromosome I between bases 2,890 and 2,991 of cosmid SPAC18G6 (EMBL, accession number Z68198). *Leu*⁺ transformants were isolated after transformation of FY1918 (*h*⁻, tRNA^{Phe}/*otr*1L(*Xho*I)::*ade*6⁺ ori1, *leu*1-32, *ura*4-D18, *ade*6-D1, *his*3-D1, *arg*3-D1) with this PCR product. PCR analyses confirmed insertion of *LEU*2 60 kb distal to

arg3. Crosses, spore viability and other procedures were as previously described³⁰.

Immunofluorescence and fluorescence *in situ* hybridization (FISH). Cells grown on malt-extract plates overnight at 32 °C were collected, washed and resuspended in 20 ml of Edinburgh minimal medium with glutamate and containing 1.2 M sorbitol. Cells were fixed for 6 min at 25 °C by the addition of paraformaldehyde to 3% and glutaraldehyde to 0.25%, and the SPB was detected by immunolocalization using affinity-purified polyclonal anti-Sad1 antibodies²¹. FISH was performed using the probe cos212 which detects the telomeres of chromosomes I and II only, as described^{4,18–20}.

Live analysis and microscopy. *Swi6*⁺ was PCR-amplified and cloned into pREP3X-GFP(S65T) (from M. Le Dizet) as a *StuI* fragment to make an N-terminal fusion with GFP. The GFP-*swi6*⁺ construct *BamHI* fragment was then subcloned into pREP42X. The resulting plasmid was linearized with *MluI*, and integrated at *arsI* on chromosome II. This was confirmed by PCR and Southern blotting. For live analyses of meiotic prophase, matings were set up between cells of opposite mating type expressing GFP-*Swi6* for both wild-type and *lot2-s17*. After incubation for 12–18 h at 32 °C, cells were scraped off and embedded in 0.6% low-melting-point agarose in minimal medium on a microscope slide, and viewed at 20–22 °C using 100× Plan Neofluar 1.3 NA objective on a Zeiss Axioplan microscope (100W Hg source) equipped with a Ludl filter wheel, and a Chroma 81000 filter set using the 81P490 filter for excitation of GFP. Digital Scientific SmartCapture software running within the IPLab host program was used to control a Photometrics Nu200 cooled CCD camera and filterwheel to obtain exposures of 1 s every 30 s. Images were stored as a multifile which could be run as an animation.

Received 11 November 1997; accepted 2 February 1998.

1. Dernberg, A. F., Sedat, J. W., Cande, W. Z. & Bass, H. W. in *Telomeres* (eds Blackburn, E. H. & Grieder, C. W.) 295–338 (Cold Spring Harbor Laboratory Press, New York, 1995).
2. Scherthan, H. *et al.* Centromere and telomere movements during early meiotic prophase of mouse and man are associated with the onset of chromosome pairing. *J. Cell Biol.* **134**, 1109–1125 (1996).
3. Bass, H. W., Marshall, W. F., Sedat, J. W., Agard, D. A. & Cande, Z. W. Telomeres cluster *de novo* before the initiation of synapsis: a three-dimensional analysis of telomere positions before and during meiotic prophase. *J. Cell Biol.* **137**, 5–18 (1997).
4. Chikashige, Y. *et al.* Telomere-led premeiotic chromosome movement in fission yeast. *Science* **264**, 270–273 (1994).
5. Scherthan, H., Bahler, J. & Kohli, J. Dynamics of chromosome organisation and pairing during meiotic prophase in fission yeast. *J. Cell Biol.* **127**, 273–285 (1994).
6. Kohli, J. Telomeres lead chromosome movement. *Curr. Biol.* **4**, 724–727 (1994).
7. Chikashige, Y. *et al.* Meiotic nuclear reorganisation: switching the position of centromeres and telomeres in the fission yeast *Schizosaccharomyces pombe*. *EMBO J.* **16**, 193–202 (1997).
8. Nimmo, E. R., Cranston, G. & Allshire, R. C. Telomere-associated chromosome breakage in fission yeast results in variegated expression of adjacent genes. *EMBO J.* **13**, 3801–3811 (1994).
9. Cooper, J. P., Nimmo, E. R., Allshire, R. C. & Cech, T. R. Regulation of telomere length and function by a Myb-domain protein in fission yeast. *Nature* **385**, 744–747 (1997).
10. Broccoli, D., Smogorzewska, A., Chong, L. & de Lange, T. Human telomeres contain two distinct Myb-related proteins TRF1 and TRF2. *Nature Genet.* **17**, 231–235 (1997).
11. Bilaud, T. *et al.* Telomeric localization of TRF2, a novel human telobox protein. *Nature Genet.* **17**, 236–239 (1997).
12. Karpen, G. H. Position-effect variegation and the new biology of heterochromatin. *Curr. Opin. Genet. Dev.* **4**, 281–291 (1994).
13. Allshire, R. C., Javerzat, J.-P., Redhead, N. J. & Cranston, G. Position effect variegation at fission yeast centromeres. *Cell* **76**, 157–160 (1994).
14. Allshire, R. C., Nimmo, E. R., Ekwall, K., Javerzat, J.-P. & Cranston, G. Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. *Genes Dev.* **9**, 218–2133 (1995).
15. Thon, G. & Klar, A. J. S. The *chl1* locus regulates the expression of the cryptic mating-type loci in fission yeast. *Genetics* **131**, 287–296 (1992).
16. Molnar, M., Bahler, J., Sipiczki, M. & Kohli, J. The *rec8* gene of *Schizosaccharomyces pombe* is involved in linear element formation, chromosome pairing and sister-chromatid cohesion during meiosis. *Genetics* **141**, 61–73 (1995).
17. Niwa, O. & Yanagida, M. Triploid meiosis and aneuploidy in *Schizosaccharomyces pombe*: an unstable aneuploid disomic for chromosome III. *Curr. Genet.* **9**, 463–470 (1985).
18. Funabiki, H., Hagan, I., Uzuwa, S. & Yanagida, M. Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast. *J. Cell Biol.* **121**, 961–976 (1993).
19. Ekwall, K. *et al.* The chromodomain protein Swi6: a key component at fission yeast centromeres. *Science* **269**, 1429–1431 (1995).
20. Ekwall, K. *et al.* Mutations in the fission yeast silencing factors *clr4*⁺ and *rik1*⁺ disrupt the localisation of the chromodomain protein Swi6p and impair centromere function. *J. Cell Sci.* **109**, 2637–2648 (1996).
21. Hagan, I. M. & Yanagida, M. The product of the spindle formation gene *sad1*⁺ associates with the fission yeast spindle pole and is essential for viability. *J. Cell Biol.* **129**, 1033–1047 (1995).
22. Heim, R., Cubitt, A. B. & Tsien, R. Y. Improved green fluorescence. *Nature* **373**, 663–664 (1995).
23. Svoboda, A., Bahler, J. & Kohli, J. Microtubule-driven nuclear movements and linear elements as meiosis specific characteristics of the fission yeasts *Schizosaccharomyces versatilis* and *Schizosaccharomyces pombe*. *Chromosoma* **104**, 203–214 (1995).
24. Munz, P. An analysis of interference in the fission yeast *Schizosaccharomyces pombe*. *Genetics* **137**, 701–707 (1994).
25. Hoheisel, J. D. *et al.* High resolution cosmid and P1 maps spanning the 14 Mb genome of the fission yeast *S. pombe*. *Cell* **73**, 109–120 (1993).
26. Mizukami, T. *et al.* A 13 kb resolution cosmid map of the 14 Mb fission yeast genome by nonrandom sequence-tagged site mapping. *Cell* **73**, 121–132 (1993).

27. Shimanuki, M. *et al.* A novel fission yeast gene, *kms1*⁺, is required for the formation of meiotic prophase specific architecture. *Mol. Gen. Genet.* **254**, 238–249 (1997).
28. Parvinen, M. & Soderstrom, K.-E. Chromosome rotation and formation of synapsis. *Nature* **260**, 534–535 (1976).
29. Lamb, N. E. *et al.* Characterization of susceptible chiasma configurations that increase the risk for maternal nondisjunction of chromosome 21. *Hum. Mol. Genet.* **6**, 1391–1399 (1997).
30. Moreno, S., Klar, A. J. S. & Nurse, P. Molecular genetic analyses of fission yeast *Schizosaccharomyces pombe*. *Meth. Enzymol.* **194**, 795–823 (1991).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

Acknowledgements. We thank K. Ekwall, J. Partridge and B. Borgstrom for discussion; and J. Cooke and N. Hastie for comments on the manuscript; M. LeDizet for pREP3X-GFP; I. Hagan for anti-Sad1 sera; K. Gould for alerting us to the potential of *his3*⁺; J. Cooper for *taz1Δ*; N. Davidson, S. Bruce and D. Stewart for photography and artwork; and A. Brown for help with the Web Page. A.L.P. is a Caledonian Research Fellow of the Royal Society of Edinburgh; E.R.N. is supported by a project grant from the Cancer Research campaign to R.C.A.

Correspondence and requests for materials should be addressed to R.C.A. (e-mail: robin.allshire@hgu.mrc.ac.uk).

Fission yeast Taz1 protein is required for meiotic telomere clustering and recombination

Julia Promisel Cooper^{*}, Yoshinori Watanabe & Paul Nurse

Cell Cycle Laboratory, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK

The alignment of homologous chromosomes during meiosis is essential for their recombination and segregation. Telomeres form and protect the ends of eukaryotic linear chromosomes, and are composed of tandem repeats of a simple DNA sequence and the proteins that bind to these repeats^{1–3}. A role for telomeres in meiosis was suspected from observations of telomere clustering in meiotic cells^{4–7}, and has now been supported experimentally by the dramatic rearrangement of telomere locations during premeiotic stages in fission yeast^{8,9}. Here we show that the fission yeast telomere protein, Taz1 (ref. 10), is required for stable association between telomeres and spindle pole bodies during meiotic prophase. In the absence of Taz1, telomere clustering at the spindle pole bodies is disrupted, meiotic recombination is reduced, and both spore viability and the ability of zygotes to re-enter mitosis are impaired to a level that would be expected if chromosome segregation were occurring randomly. Such telomeric association mediated by telomere-specific proteins may also be important for proper chromosome alignment and recombination during meiosis in humans.

The fission yeast protein Taz1 organizes a telomere-specific nucleoprotein structure¹⁰ (J.P.C. and T. R. Cech, manuscript in preparation). Deletion of *taz1* results in substantial elongation of the telomeric DNA tract, disappearance of telomere-specific chromatin structure, and loss of the repression of telomere-adjacent gene expression known as the telomere position effect^{10–12}. Vegetative growth is largely unimpaired in *taz1*[−] cells, which indicates that telomere length restriction and telomere position effects are largely dispensable for viability. Sexual reproduction is impaired, however, with the formation of aberrant tetrads whose spores are predominantly inviable¹⁰.

During premeiotic stages in wild-type cells, telomeres become localized next to the spindle pole bodies (SPBs) at the leading edge of chromosomal movements^{8,9}. To determine whether the attachment

^{*} Present address: Department of Biochemistry and Molecular Genetics, University of Colorado Health Sciences Center, Denver, Colorado 80262, USA.

of telomeres to SPBs is disrupted in *taz1*⁻ cells, we examined chromosomal locations by fluorescence *in situ* hybridization (FISH; Figs 1, 2). Nuclei in the 'horsetail' stage of meiotic prophase (Fig. 3) were scored, because their ready identification prevents any ambiguities that might arise from comparing cells in different stages of meiosis. In wild-type crosses, all telomeres on chromosomes I, II (Fig. 1a) and III (Fig. 1b) cluster together in a single spot which localizes to the SPB^{8,9} (Fig. 2a). In contrast, the majority of *taz1*⁻ horsetail nuclei show more than one spot of hybridization to telomere-adjacent probes, indicating that Taz1 is required to hold telomeres in clusters. Concomitant with this dispersal of the telomere signal, 79% of *taz1*⁻ horsetails contain one or more telomere spot(s) that venture far from the SPB, and many (~31%) have a complete lack of telomere-SPB association. Thus, it appears that the affinities governing associations between telomeres and SPBs are lowered considerably in the absence of *taz1*. Centromere locations are not detectably affected by *taz1* deletion as centromeres are displaced from SPBs in both wild-type and *taz1*⁻ horsetail nuclei (Fig. 1c).

The appearance of the horsetail nucleus itself is also altered in many *taz1*⁻ cells (see DAPI panels in Fig. 1). Instead of the smoothly elongated horsetails seen in wild-type cells, *taz1*⁻ horsetails often appear lumpy or disjointed, and the region of the horsetail containing the SPB often appears to be connected to the bulk of the nuclear material by a thin thread. If the characteristic horsetail shape in wild-type cells stems from a lengthwise alignment of the chromosomes that is achieved through their clustering at SPBs, it is perhaps not surprising that the horsetail shape is distorted when clustering and alignment are impaired.

Disruption of telomeric clustering at SPBs by the *taz1* deletion allows an investigation of the role of such attachments in meiotic processes. We first asked whether intragenic recombination is impaired during a *taz1*⁻ meiosis, using *ade6-M26* and *ade6-469* alleles as markers; recombination between the two defective alleles yields a functional product, conferring growth on medium lacking adenine¹³. The results show a reduction of recombination in the *taz1*⁻ background (Fig. 2b); the percentage of viable spores that had undergone recombination between *ade6-M26* and *ade6-469* was reduced in *taz1*⁻ crosses by ~3.5-fold relative to wild type. Intergenic recombination between the *leu1* and *mat1* loci is also greatly reduced in *taz1*⁻ cells, with recombination between these loci dropping from 22% in wild-type crosses to 2% in *taz1*⁻ crosses.

Taz1 is therefore required for efficient meiotic recombination. This requirement is most likely to stem from the role of Taz1 in gathering telomeres at SPBs during meiotic prophase, thereby initiating chromosomal alignment and promoting homologous recombination. Given that spore viability is only ~7% (ref. 10; Fig. 1c), the true effect of *taz1* deletion on recombination may be greater, but because it often leads to a complete failure of sporulation, this effect cannot be properly quantified. *kms1*⁻ strains also produce abnormal asci and have a reduced meiotic recombination frequency, although they are recessive¹⁴. *kms1*⁻ cells show fragmentation of both telomeric and SPB signals during meiotic prophase; in contrast, most *taz1*⁻ horsetail nuclei exhibit a single SPB signal from which the telomere signals often stray. Hence, Kms1 may represent a SPB component that does not interact directly with

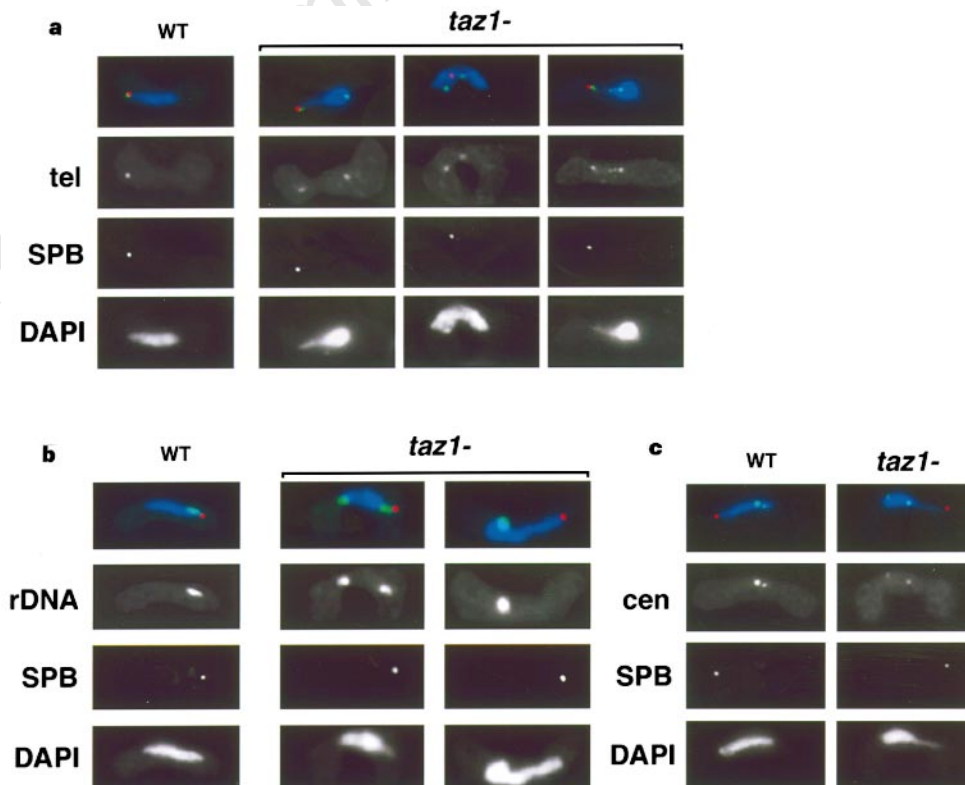


Figure 1 Locations of telomeres and centromeres during meiotic prophase in wild-type and *taz1*⁻ cells. Homothallic cells were grown to early log phase and then induced to initiate meiosis by transferring to sporulation medium (EMM medium, without nitrogen) for 10–12 hours before fixation. Fixed cells were subjected to indirect immunofluorescence using an antibody against the SPB (anti-Sad1; refs 22, 24; red spots), refixed, and subjected to FISH. DAPI staining of nuclear chromatin is shown in blue. **a**, FISH was performed using the telomere-

associated sequence probe²² (tel; green spots in top panels), which hybridizes to the telomeres on either end of chromosomes I and II. **b**, FISH was performed using a ribosomal DNA probe²⁶ (rDNA; green spots in top panel). The rDNA region is located adjacent to either telomere on chromosome III, and spans a total of ~1 megabase. **c**, FISH was performed using a centromere probe²⁶ (cen; green spots in top panel).

telomeres, whereas Taz1 participates in the connection between telomeres and SPBs but is not essential for the integrity of the SPB itself.

As Taz1 is required for progression through meiosis, we investigated the meiotic events in which Taz1 participates. *taz1*⁻ haploids

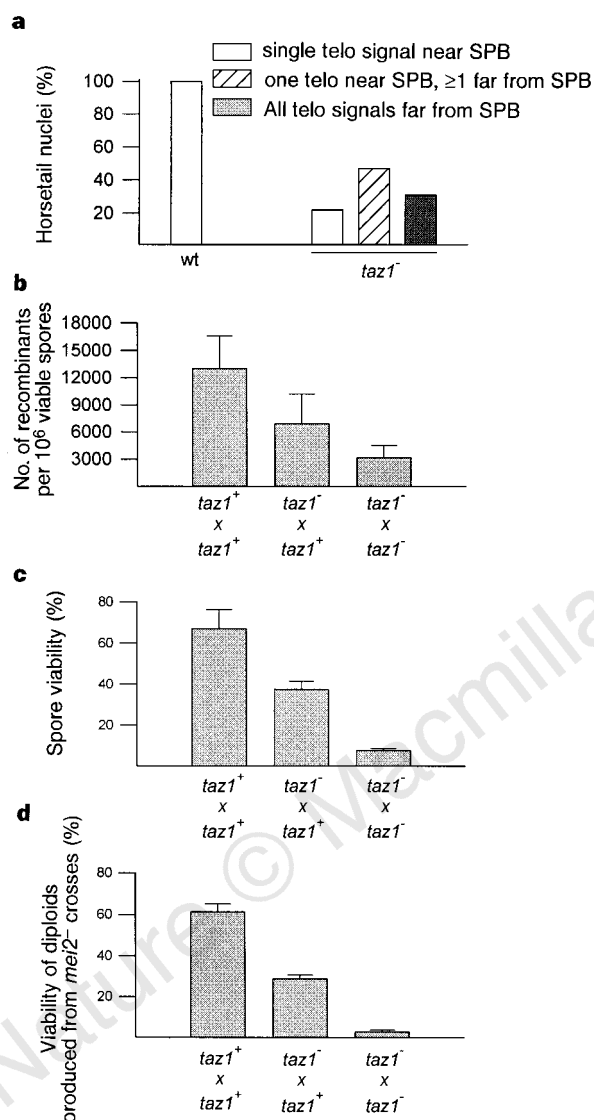


Figure 2 Effects of *taz1* deletion on telomere clustering, meiotic recombination, spore viability and resumption of mitotic growth. Each bar represents four (b), six (c) or two (d) independent experiments, with error bars indicating standard deviations. **a**, Telomere clustering at SPBs is disrupted in *taz1*⁻ horsetail nuclei. Sample preparation and probes are described in Fig. 1. The graph represents scoring of SPB and rDNA signals in 81 horsetail nuclei. Virtually identical results were obtained from scoring SPB and tel signals. **b**, Intragenic meiotic recombination is reduced in *taz1*⁻ crosses. *ade6-M26* and *ade6-469* mutants were crossed with the strains indicated, mated on SPA medium for 48 h, and plated either on EMM with adenine to determine total spore viability, or on EMM without adenine to determine the percentage of viable spores that had undergone recombination between *ade6-M26* and *ade6-469*. **c**, Crosses of *taz1*⁻ haploids lead to reduced spore viability. Log phase cultures of the strains indicated were mixed on SPA medium, incubated overnight at 29°C, treated with helicase, and plated on supplemented EMM medium. **d**, *taz1* deletion reduces the ability of cells arrested in meiosis to resume mitotic growth. *mei2* cells were crossed with the indicated strains and mated overnight on SPA. Zygotes were counted and plated on EMM (minus adenine).

mate to form zygotes with about the same efficiency as wild-type haploids (wild-type crosses yield ~58% zygotes, *taz1*⁻ crosses yield ~66%), suggesting that telomere attachment to SPBs is not required for the conjugation process itself. We have also noticed that heterozygous crosses between *taz1*⁺ and *taz1*⁻ haploids lead to defective sporulation, with the absence of *taz1* in one mating partner reducing spore viability to a level intermediate between that obtained from homozygous *taz1*⁺ and *taz1*⁻ crosses (Fig. 1c). Of the viable spores produced from *taz1*⁺ and *taz1*⁻ parents, about equal numbers are *taz1*⁺ and *taz1*⁻, indicating that deletion of *taz1* confers no selective disadvantage with regard to spore germination or subsequent mitotic growth. In contrast to the defective sporulation stemming from a *taz1*⁺ × *taz1*⁻ mating, no sporulation defects are apparent when a vegetatively growing heterozygous diploid is induced to undergo azygotic meiosis (data not shown). The *taz1*⁻ meiotic defect therefore originates in haploid mating partners before conjugation, consistent with results⁹ showing that the attachment of telomeres to SPBs is initiated in wild-type haploid cells upon exposure to mating pheromone and persists throughout meiotic prophase (Fig. 3).

We next investigated whether *taz1*⁻ zygotes can return to mitotic growth after initiating meiosis. For these experiments we exploited the ability of the mutant strain *mei2*⁻ to resume mitotic growth after having mated and proceeded to an early stage of meiosis; *mei2*⁻ cells can conjugate but arrest before premeiotic DNA synthesis^{15,16}. The

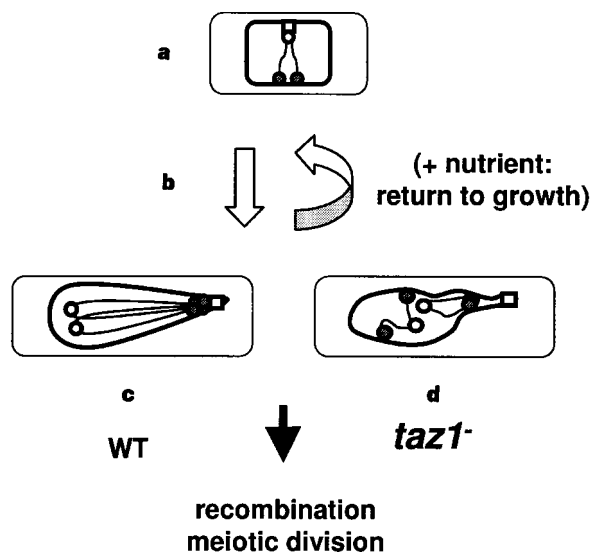


Figure 3 Taz1 is required for shifts in telomere locations during meiosis in fission yeast. White squares, SPBs; grey circles, telomeres; white circles, centromeres. **a**, Before induction of meiosis, centromeres are located adjacent to SPBs and telomeres are located in several positions near the nuclear periphery. **b**, Upon exposure of haploid cells to compatible mating pheromones, all telomeres (there are 6 on the 3 fission-yeast chromosomes; only one chromosome is represented here for each 1C DNA content) become localized in a single cluster adjacent to the SPB and the centromeres also remain there⁹. During conjugation and nuclear fusion, telomeres remain close to SPBs and centromeres are no longer located there^{8,9}. **c**, The nucleus assumes the so-called horsetail configuration and moves back and forth across the cell. During this period, telomeres remain close to the SPBs at the leading edge of chromosomal movement; this is the period during which meiotic prophase occurs. The nucleus returns to the centre of the cell before meiosis I, when centromeres resume their SPB-adjacent positions and telomeres become separated from the SPBs⁹. **d**, In *taz1*⁻ cells, centromeres become detached from SPBs but telomeres do not form stable associations (either with each other or with SPBs). Chromosomal association, recombination and segregation are impaired, leading to low spore viability, as well as a reduced ability to return to mitotic growth once meiosis has been initiated.

ability to return to vegetative growth was assessed by mating the two complementing alleles *ade6-M210* and *ade6-M216*, because neither *ade6-M210* nor *ade6-M216* haploids can survive on media lacking adenine, whereas diploids containing both alleles can grow¹⁷. Thus if *ade6-M210* and *ade6-M216* strains are induced to undergo meiosis and then transferred to medium lacking adenine, the haploid offspring of this pairing will be unable to grow as none will contain both *ade6* alleles. However, if a *mei2* deletion strain is subjected to the same treatment, its inability to undergo meiosis results in cells arresting as *ade6-M210/ade6-M216* zygotes which are viable as diploids on medium lacking adenine.

Deletion of *taz1* in the *mei2*⁻ background causes a marked reduction in the ability of zygotes to return to mitotic growth (Fig. 2d). The reduction in viability of the diploid cells upon return to mitotic growth shows similarities with the reduction in spore viability seen in *taz1*⁻*mei2*⁺ strains (Fig. 2c), with the heterozygous *taz1*⁻*mei2*⁻ × *taz1*⁺*mei2*⁻ cross showing intermediate diploid viability. This suggests that the resumption of vegetative growth from an early stage of meiosis is subject to the same limitation as is the completion of meiosis to form viable spores. Once centromeres have become detached from SPBs during conjugation, *taz1*⁻ zygotes, in which the attachments between telomeres and SPBs are tenuous or absent, are unable to align chromosomes properly for the first nuclear division, whether it be meiotic or mitotic. Indeed, the levels of both spore viability (Fig. 2c) and diploid viability (Fig. 2d) correspond approximately to the levels that would be expected if the three fission yeast chromosomes were segregating randomly, having no stable points of SPB attachment for orientation.

Our results suggest that the absence of a single protein, Taz1, results in the destabilization of telomere clustering at SPBs and the consequent disarray of chromosomal organization during meiotic prophase (Fig. 3). This in turn impinges on the ability of chromosomes to align properly for meiotic recombination and to segregate properly during meiosis. Our observation that the *taz1* deletion severely reduces the ability of *mei2* mutants to resume mitotic growth indicates that premeiotic chromosomal rearrangements become critical for subsequent cell divisions, even before premeiotic DNA synthesis has occurred. Telomere clustering during meiosis has been observed in many eukaryotes⁴⁻⁷, and the association of telomeres mediated by telomere-specific proteins may therefore be important for chromosome alignment and recombination in eukaryotes other than fission yeast. The human protein hTRF1 (ref. 18) associates with telomeres and shares significant structural and functional homology with Taz1 (refs 18, 19), suggesting that such mechanisms may operate in human cells. □

Methods

Media and yeast. Standard methods were used²⁰, with YE as complete medium and Edinburgh minimal medium or synthetic sporulation medium²¹ as indicated. Cells were incubated at 32°C for vegetative growth or at 29°C to induce meiosis. Strains contained only the markers indicated in the figure legends, with the exception of Figs 1, 2a and c, which used strains harbouring the additional markers *leu1-32* and *ade6-M210* or *ade6-M216*. The *taz1*⁻ deletion lacks the entire *taz1* reading frame (*taz1::ura4*⁺)¹⁰; the *mei2*⁻ deletion lacks the entire *mei2* reading frame (*mei2::ura4*⁺)¹⁶.

Fluorescence in situ hybridization. Cells induced for meiosis were collected and fixed in PEM with 3% paraformaldehyde (pFA) and 0.2% glutaraldehyde for 10 min at room temperature. Following fixation, cells were washed, digested with Zymolyase and Novozyme, permeabilized and treated with RNaseA as described^{18,9,22,23}. SPBs were detected by indirect immunofluorescence with anti-Sad1 antibody²⁴ and a Cy3-conjugated second antibody. Cells were then refixed in PEM with 3% pFA and 0.2% glutaraldehyde for 1 h at room temperature before denaturation and hybridization as described^{8,23}. cos212 (ref. 22), Yip10.4 (ref. 25) and pRS140 (ref. 26) were used as hybridization probes for telomere-associated sequences, rRNA genes and centromeres, respectively. Probes were labelled with digoxigenin by using a nick-translation kit (Boehringer Mannheim), purified over Qiagen PCR purification columns, and detected

with a fluorescein-conjugated anti-digoxigenin antibody. Images were collected on a Zeiss axioplan microscope by using either a Princeton Instruments cooled CCD camera or a Hamamatsu chilled video-rate CCD camera.

Received 3 November 1997; accepted 18 February 1998.

- Blackburn, E. H. & Greider, C. W. *Telomeres* (Cold Spring Harbor Laboratory Press, New York, 1995).
- Greider, C. W. & Blackburn, E. H. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* **43**, 405–413 (1985).
- Zakian, V. A. Telomeres: beginning to understand the end. *Science* **270**, 1601–1607 (1995).
- Loidl, J. The initiation of meiotic pairing: the cytological view. *Genome* **33**, 759–778 (1990).
- John, B. Meiosis. *Dev. Cell Biol. Ser.* **22** (Cambridge University Press, Cambridge, 1990).
- Scherthan, H., Bahler, J. & Kohli, J. Dynamics of chromosome organization and pairing during meiotic prophase in fission yeast. *J. Cell Biol.* **127**, 273–285 (1994).
- Dernburg, A. F., Sedat, J. W., Cande, W. Z. & Bass, H. W. in *Telomeres* 295–338 (Cold Spring Harbor Laboratory Press, New York, 1995).
- Chikashige, Y. et al. Telomere-led premeiotic chromosome movement in fission yeast. *Science* **264**, 270–273 (1994).
- Chikashige, Y. et al. Meiotic nuclear reorganization: switching the position of centromeres and telomeres in the fission yeast *Schizosaccharomyces pombe*. *EMBO J.* **16**, 193–202 (1997).
- Cooper, J. P., Nimmo, E. R., Allshire, R. C. & Cech, T. R. Regulation of telomere length and function by a Myb-domain protein in fission yeast. *Nature* **385**, 744–747 (1997).
- Gottschling, D. E., Aparicio, O. M., Billington, B. L. & Zakian, V. A. Position effect at *Saccharomyces cerevisiae* telomeres: reversible repression of pol II transcription. *Cell* **63**, 751–762 (1990).
- Nimmo, E. R., Cranston, G. & Allshire, R. C. Telomere-associated chromosome breakage in fission yeast results in variegated expression of adjacent genes. *EMBO J.* **13**, 3801–3811 (1994).
- Bahler, J., Schuchert, P., Grimm, C. & Kohli, J. Synchronized meiosis and recombination in fission yeast: observations with *pat1-114* diploid cells. *Curr. Genet.* **19**, 445–451 (1991).
- Shimanuki, M. et al. A novel fission yeast gene, *kms1*⁺, is required for the formation of meiotic prophase-specific nuclear architecture. *Mol. Gen. Genet.* **254**, 238–249 (1997).
- Egel, R. Commitment to meiosis in fission yeast. *Mol. Gen. Genet.* **121**, 277–285 (1973).
- Watanabe, Y. & Yamamoto, M. *S. pombe mei2*⁻ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species *meiRNA*. *Cell* **78**, 487–498 (1994).
- Gutz, H. Site-specific induction of gene conversion in *Schizosaccharomyces pombe*. *Genetics* **69**, 317–337 (1971).
- Chong, L. et al. A human telomeric protein. *Science* **270**, 1663–1667 (1995).
- van Steensel, B. & de Lange, T. Control of telomere length by the human telomeric protein TRF1. *Nature* **385**, 740–743 (1997).
- Moreno, S., Klar, A. J. S. & Nurse, P. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Meth. Enzymol.* **194**, 795–823 (1991).
- Gutz, H., Heslot, H., Leupold, U. & Loprieno, N. in *Handbook of Genetics* 211–287 (Plenum, New York, 1974).
- Funabiki, H., Hagan, I., Uzawa, S. & Yanagida, M. Cell cycle specific positioning and clustering of centromeres and telomeres in fission yeast. *J. Cell Biol.* **121**, 961–976 (1993).
- Uzawa, S. & Yanagida, M. Visualization of centromeric and nucleolar DNA in fission yeast by fluorescence *in situ* hybridization. *J. Cell Sci.* **101**, 267–275 (1992).
- Hagan, I. & Yanagida, M. The product of the spindle formation gene *sad1*⁺ associates with the fission yeast spindle pole body and is essential for viability. *J. Cell Biol.* **129**, 1033–1047 (1995).
- Toda, T., Nakaseko, Y., Niwa, O. & Yanagida, M. Mapping of rRNA genes by integration of hybrid plasmids in *Schizosaccharomyces pombe*. *Curr. Genet.* **8**, 93–97 (1984).
- Chikashige, Y. et al. Composite motifs and repeat symmetry in *Schizosaccharomyces pombe* centromeres: direct analysis by integration of *NotI* restriction sites. *Cell* **57**, 739–751 (1989).

Acknowledgements. We thank all the members of our laboratory for discussion, and particularly J. Hayles, J. Mata and K. Sawin for help and advice, K. Ekwall for cosmid, and I. Hagan for anti-Sad1 antibody. J.P.C. was supported by an EMBO fellowship and Y. W. was supported by a JSPS fellowship.

Correspondence and requests for materials should be addressed to J.P.C. (e-mail: cooperj@essex.uchsc.edu).

Transcriptional repression by UME6 involves deacetylation of lysine 5 of histone H4 by RPD3

Stephen E. Rundlett*, Andrew A. Carmen*, Noriyuki Suka*, Bryan M. Turner† & Michael Grunstein*

* Department of Biological Chemistry, UCLA School of Medicine, Los Angeles, California 90095, USA

† The University of Birmingham, Chromatin and Gene Expression Group, Department of Anatomy, The Medical School, Birmingham B15 2TT, UK

The histone deacetylase RPD3 can be targeted to certain genes through its interaction with DNA-binding regulatory proteins. RPD3 can then repress gene transcription¹⁻⁶. In the yeast *Saccharomyces cerevisiae*, association of RPD3 with the transcriptional repressors SIN3 and UME6 results in repression of reporter genes containing the UME6-binding site³. RPD3 can deacetylate

all histone H4 acetylation sites in cell extracts⁷. However, it is unknown how H4 proteins located at genes near UME6-binding sites are affected, nor whether the effect of RPD3 is localized to the promoter regions. Here we study the mechanism by which RPD3 represses gene activity by examining the acetylation state of histone proteins at UME6-regulated genes. We used antibodies specific for individual acetylation sites in H4 to immunoprecipitate chromatin fragments. A deletion of *RPD3* or *SIN3*, but not of the related histone-deacetylase gene *HDA1*, results in increased acetylation of the lysine 5 residue of H4 in the promoters of the UME6-regulated *INO1* (ref. 8), *IME2* (ref. 3) and *SPO13* (ref. 9) genes. As increased acetylation of this residue is not merely a consequence of gene transcription, acetylation of this site may be essential for regulating gene activity.

To determine the extent to which mutant histone deacetylases affect transcription of UME6-regulated genes in our yeast strains, we examined the levels of messenger RNA of several genes through northern blotting of polyadenylated RNA from wild-type cells and *hda1* and *rpd3* mutant cells (Fig. 1). Unlike *hda1*, the *rpd3* mutation leads to a 32-fold and 3.6-fold increases in the expression of *INO1* and *IME2*, respectively. *SPO13* expression was not increased in our strains. For comparison, we also examined the effects of histone deacetylase mutations on the transcription of control genes (*PHO3*, *SPS2*, *INH1*, *ERG10* and *SIP2*) not known to be affected by UME6. Although transcription of *SPS2* and *SIP2* was not altered appreciably, *INH1* and *ERG10* mRNA levels were similarly increased by both deacetylase mutations. *PHO3* mRNA levels were increased more by *rpd3* (3.7-fold increase) than by *hda1* (2.5-fold increase). Thus, a deletion in *RPD3* leads to increased expression of several genes, including UME6-regulated *INO1* and *IME2*.

We next investigated whether increased acetylation of histone H4 accompanies the increased transcription of these genes. Using antibodies that recognize acetylated lysine at histone H4 positions K5, K8, K12 or K16 specifically, we immunoprecipitated formaldehyde-crosslinked, sonicated chromatin fragments from wild-type

and mutant cells and analysed the immunoprecipitated DNA with the polymerase chain reaction (PCR). We designed primer sets to include upstream regulatory sequences of the genes mentioned above. A similar procedure has been used previously to localize proteins at telomeric heterochromatin¹⁰. If the deacetylase mutation resulted in greater representation of certain genes within immunoprecipitated DNA, we took this to mean that the gene fragment in question is hyperacetylated. We also examined the acetylation state of a region 0.5 kilobases (kb) from the telomere of chromosome VI-R; this is a site that is not transcribed and which serves as a reference for loading differences. We found that the *rpd3* mutation (Fig. 2a, lane 3) leads to increased acetylation of H4 K5 within the promoters of *INO1* (~5-fold increase), *IME2* (~5-fold increase) and *SPO13* (~3-fold increase), as shown by quantification (Fig. 2c) of the increased intensity of these PCR products compared with either wild-type (Fig. 2a, b, lane 1) or *hda1* mutant (Fig. 2a, b, lane 2) cells. These increases contrast with the relatively minor changes in acetylation at K8 (Fig. 2a, b, lanes 4–6), K12 (Fig. 2a, b, lanes 7–9) and K16 (Fig. 2a, b, lanes 10–12). *RPD3* deletion does lead to a small increase (~1.5-fold increase) in K5 acetylation at the *PHO3* promoter but does not seem to affect *SPS2*, *INH1*, *ERG10*, and *SIP2*, appreciably. We conclude that *RPD3* (but not *HDA1*) affects

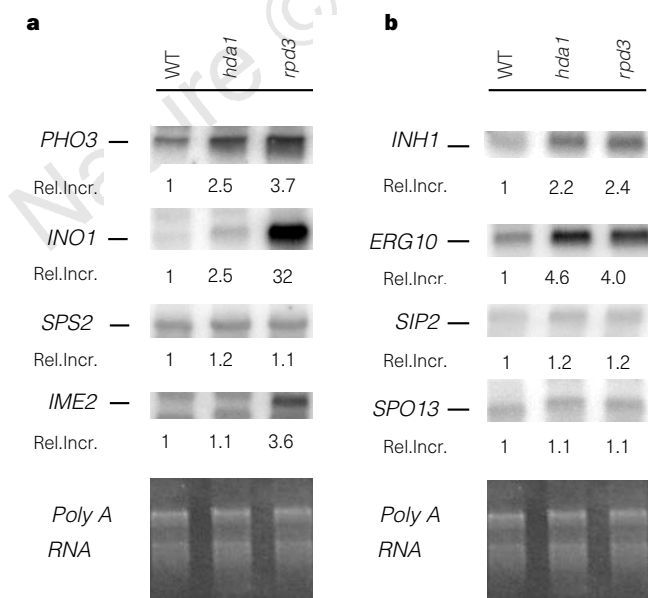


Figure 1 Effect of *RPD3* and *HDA1* disruption on gene expression. Polyadenylated RNA from either wild-type (WT), *hda1* or *rpd3* cells was hybridized by northern blotting to probes specific for either **a**, the *PHO3*, *INO1*, *SPS2* and *IME2* genes or **b**, the *INH1*, *ERG10*, *SIP2* and *SPO13* genes. The numbers are the quantified relative increase in mRNA level (Rel. Incr.) in mutant compared with WT strains. Photographs of the ethidium-bromide-stained gels before transfer are shown at the bottom to illustrate amounts of poly(A) RNA loaded.

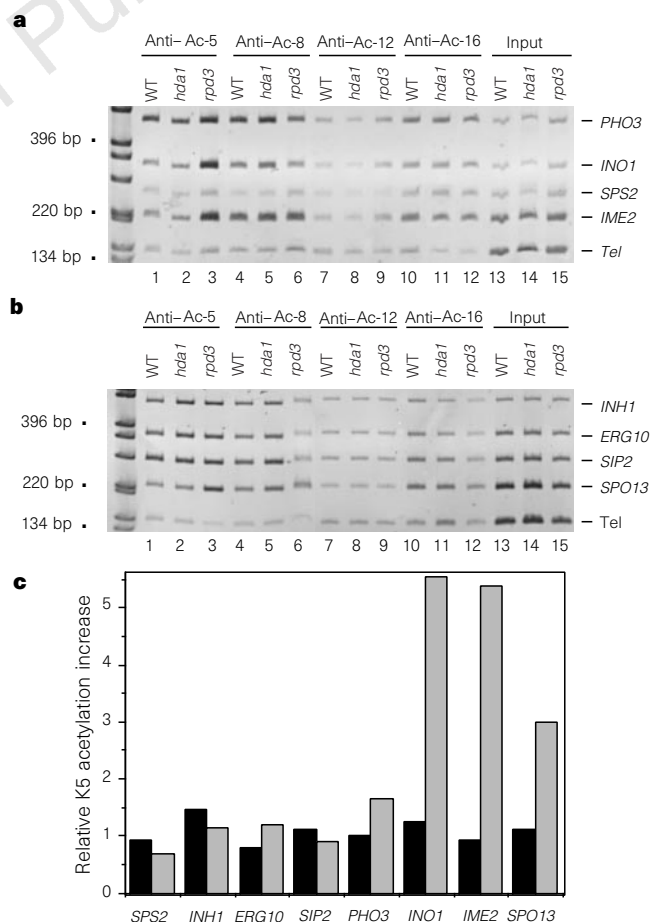


Figure 2 Effect of *RPD3* disruption on acetylation of histone H4 at promoters of specific genes. **a**, **b**, Multiplex PCR analysis of DNA immunoprecipitated by antibodies against H4 acetylated lysines 5, 8, 12 or 16. DNA from wild-type (WT), *hda1* and *rpd3* strains and input DNA obtained before immunoprecipitation were used. Primers were used which span promoter regions of **a**, *PHO3*, *INO1*, *SPS2* and *IME2* and **b**, *INH1*, *ERG10*, *SIP2* and *SPO13*. Amplification of a fragment 0.5 kb from the telomere (Tel) of chromosome VI-R was used as reference to ensure equal loading of samples. **c**, Quantification of increases in H4 K5 acetylation in *hda1* (black columns) and *rpd3* (grey columns) cells relative to wild-type cells.

H4 K5 acetylation most strongly within the promoters of the *INO1*, *IME2* and *SPO13* genes.

RPD3 is part of a complex of high relative molecular mass that includes SIN3 (refs 11, 12). Therefore, we determined whether a *SIN3* deletion (*sin3*) would also lead to increased histone acetylation. The *sin3* mutation leads to an increase in acetylation of H4 K5 at *INO1* (3.6-fold increase), *IME2* (3.2-fold increase) and *SPO13* (3.3-fold increase) (Fig. 3a, b, lanes 1, 2). There is no obvious effect on H4 K8 (Fig. 3a, b, lanes 3, 4) or K16 (Fig. 3a, b, lanes 7, 8). In contrast, the *sin3* mutation leads to a ~3-fold increase in acetylation at H4 K12 (Fig. 3a, b, lanes 5, 6) at *INO1*, *IME2* and *SPO13*. Interestingly, there is also an increase in K5 and K12 acetylation at *PHO3*. We conclude that SIN3 regulates acetylation of K5 at UME6-regulated genes, in a similar manner to RPD3, indicating that RPD3 and SIN3 may be members of the same complex. However, as the *sin3* mutation also leads to K12 hyperacetylation at some loci, SIN3 may affect the specificity of deacetylation by RPD3 or interact with other deacetylases.

UME6-binding sites are located in regulatory regions upstream of the *INO1*, *IME2*, and *SPO13* coding regions^{8,13}. One such site (upstream regulatory site (URS)1) at *INO1* mediates repression by SIN3 when fused to a reporter gene⁸. To investigate whether RPD3 deacetylates H4 K5 near the regulatory region or throughout the entire gene, we first used primer sets for PCR that were located at

roughly 1-kb intervals across the *INO1* gene (Fig. 4a). As nucleosomes are often excluded from DNA sites that bind regulatory proteins¹⁴, we used a promoter primer set in which the 5' primer of the pair was adjacent to the URS. The precise location (−0.09 kb) of the primers is also determined by the need for an ~50% G + C content¹⁰ and a size that allows its distinction from other final PCR-generated fragments in the multiplex PCR analysis. Our analysis shows that in both wild-type (Fig. 4b, lane 1) and *hda1* mutant (Fig. 4b, lane 2) cells, the extent of K5 acetylation at the promoter (−0.09 kb) is roughly threefold lower than at distances approximately 1 kb (−1.1 kb and +1.0 kb) or 2 kb (−2.0 kb and +1.9 kb) away. Thus, the promoter-adjacent region is specifically hypoacetylated at K5 in wild-type cells (Fig. 4b, lane 3) and is then strongly acetylated when *RPD3* is deleted.

To more precisely investigate the distance from the promoter-adjacent region at which RPD3 modifies nucleosomes, we used primers in which the 5' ends of the products were within −0.73 kb and +0.14 kb of the translation start site. As chromatin is randomly sheered to an average length of ~0.5 kb, with a range of 0.3–0.8 kb, before immunoprecipitation (data not shown), the distances of the PCR products from the promoter (−0.09 kb) fragment are near the limit of resolution for this technique. Through the use of a primer set located 1.9 kb upstream of *INO1* as an internal control for the PCR, we found that only the region at the promoter

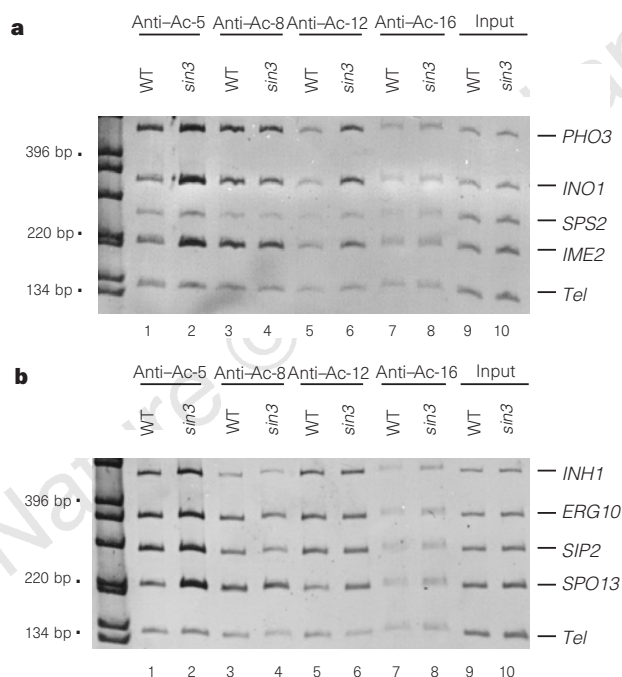


Figure 3 Effect of *SIN3* disruption on acetylation of specific genes. Multiplex PCR analysis was performed on DNA immunoprecipitated by antibodies against histone H4 acetylated lysines 5, 8, 12 or 16. DNA from wild-type (WT) and *sin3* cells and input DNA were used. PCR amplification was done with the primer sets spanning the promoters of **a**, *PHO3*, *INO1*, *SPS2* and *IME2* and **b**, *INH1*, *ERG10*, *SIP2* and *SPO13*. The telomere (Tel) DNA control was used as in Fig. 2.

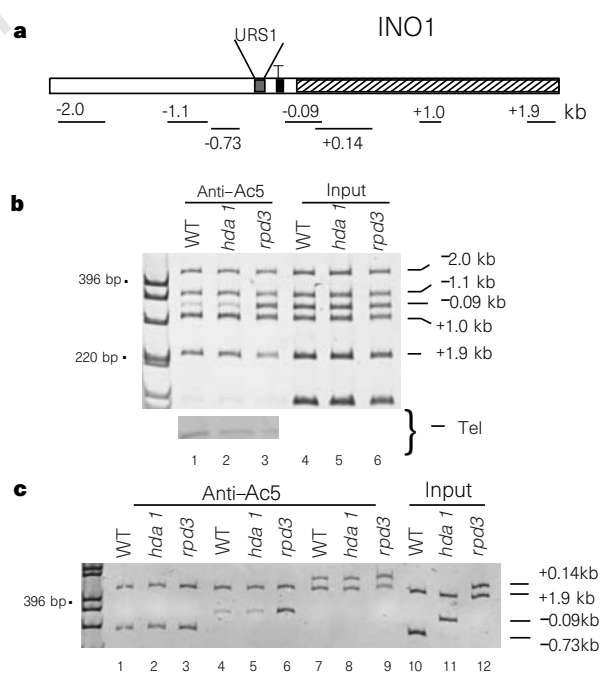


Figure 4 Effect of *RPD3* disruption on histone acetylation next to the UME6-binding site (URS1). **a**, The distance from the start of translation (ATG) for the 5' end of each PCR product is shown in kilobases (kb). The ATG used for translation is derived from sequence databases and the primary sequence of the *INO1* protein¹⁶. The TATA box (T) and the coding region (hatched area) of *INO1* are also shown. **b**, Chromatin was immunoprecipitated using an antibody against H4 acetylated lysine 5 antibody and DNA from wild-type (WT) (lane 1), *hda1* (lane 2) and *rpD3* (lane 3) cells was analysed by PCR using primer sets located at roughly 1-kb intervals across the *INO1* gene. WT, *hda1* and *rpD3* input DNAs are shown in lanes 4, 5 and 6. The Tel row represents an overexposure of the hypoacetylated telomere region, to show similar loading. **c**, Higher resolution chromatin immunoprecipitation analysis of the region next to the URS1. PCR primers were used at a distance of −0.73 kb upstream of the translation start (lanes 1–3), at the promoter (−0.09 kb) (lanes 4–6), and +0.14 kb downstream of the translation start (lanes 7–9). A primer set +1.9 kb upstream of the gene was used as an internal control for each PCR reaction.

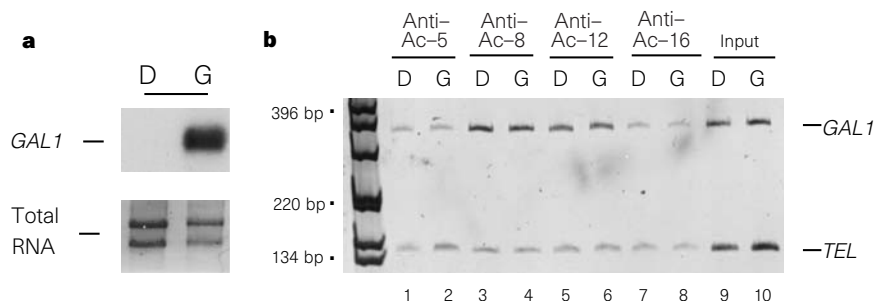


Figure 5 Effect of galactose induction on transcription and H4 acetylation of *GAL1*. **a**, Total RNA was prepared from YDS3 cells grown in glucose (YEPD medium; D) or galactose media (YEPG medium, G)^{17,18} and was electrophoresed, blotted and probed with a *GAL1*-containing *SalI/EcoRI* fragment of plasmid pNN78 (ref. 19). **b**, YDS3 cells were grown in either glucose (D) or galactose (G)

and chromatin was immunoprecipitated using antibodies against histone H4 acetylated lysines 5, 8, 12 or 16 as described in Fig. 2. Input DNAs are also shown. The immunoprecipitated DNA was analysed by PCR using primers that amplified the *GAL1* promoter. A fragment 0.5 kb from the telomere of chromosome VI-R was used as a reference.

(−0.09 kb) is hypoacetylated (Fig. 4c, lanes 4–6) and is then acetylated when RPD3 is deleted. Regions at either −0.73 kb upstream (lanes 1–3) or +0.14 kb downstream (Fig. 4c, lanes 7–9) are relatively unaffected by RPD3 disruption. Therefore, the region affected by RPD3 is probably between the UME6-binding site and the upstream boundary of the primer set at +0.14, a distance of 386 base pairs (bp). As nucleosome and linker DNA measure ~165 bp in yeast¹⁴, RPD3 is likely to deacetylate H4 at two nucleosomes or less from the UME6-binding site.

Although there was an increase in *INH1* and *ERG10* gene expression in *rp3* mutant strains (Fig. 1b), there was relatively little alteration in histone H4 acetylation at these genes (Fig. 2b), indicating that the process of transcription itself probably does not lead to histone acetylation. Conversely, there is an increase in K5 acetylation at *SPO13*, which is UME6-regulated, but there is no obvious increase in *SPO13* mRNA levels in the *rp3* mutant background. However, our northern analysis may be too insensitive to detect subtler increases in transcription. To rule out the possibility that a dramatic increase in mRNA levels, such as the increase in *INO1* mRNA levels, could alter acetylation of H4 we determined whether increased expression of the *GAL1* gene could lead to increased acetylation of H4 near the *GAL1* promoter. Wild-type cells were grown in either repressive (2% glucose) or induced (2% galactose) conditions; we then immunoprecipitated their chromatin and used it for PCR analysis to investigate acetylation at the *GAL1* promoter. Despite the large induction of *GAL1* transcription (Fig. 5a), there is no evident change in acetylation of histone H4 K5 (Fig. 5b, lanes 1, 2), K8 (Fig. 5b, lanes 3, 4), K12 (Fig. 5b, lanes 5, 6) or K16 (Fig. 5b, lanes 7, 8) within the *GAL1* promoter. Thus, increased transcription alone does not increase histone H4 acetylation.

Our results indicate that localization of the histone deacetylase RPD3 and the co-repressor SIN3 at UME6-regulated genes deacetylates adjacent histone H4 K5 and represses transcription. It is likely that sites in other histones are also affected by RPD3 as the use of an antibody that recognizes both H3 acetylated lysines, K9 and K18, does show some increase in acetylation at the *INO1*, *IME2* and *SPO13* genes in *rp3* but not *hda1* mutant cells (data not shown). For more detailed analysis, antibodies against the remaining individual sites of acetylation in the core histones must be produced. Nevertheless, as the *rp3* mutation results in increased acetylation of all four sites of acetylation in H4 in cell extracts⁷, chromatin or other components of the deacetylation complex may limit the specificity of RPD3 to H4 K5 *in vivo*. SIN3 may have such a role, as its disruption results in increased K5 and K12 acetylation. However, SIN3 may have functions in repression that are independent of RPD3. For example, there is still some residual repression through SIN3 when RPD3 is deleted¹. The possibility has not been excluded

that SIN3 may recruit a second histone deacetylase (such as HDA1, HOS1, HOS2 or HOS3 (ref. 7)) to UME6-regulated genes to deacetylate K12.

Only K5 of the H4 sites at the *INO1* promoter may be hypoacetylated (Fig. 2). Acetylation of K5 would then lead to the tetra-acetylated form of H4. Recent X-ray crystallographic data have shown that residues 16–25 of H4 interact with the histone H2A–H2B dimer in an adjacent nucleosome of the crystal lattice¹⁵. If interactions are similar in chromatin, the acetyltable H4 tail may link adjacent nucleosomes. Thus, the deacetylated state of K5 may promote nucleosome–nucleosome packing in regulatory elements, thereby limiting gene activity. □

Methods

mRNA levels. Purified poly(A) RNA was isolated from either YDS3 (wild-type)²⁰, SRY3D2 (*hda1*) or SRYR35 (*rp3*)⁷ cells through the use of a hot-phenol-based techniques¹⁷. All cells were grown to $A_{600} \sim 1.5$ in YEPD medium¹⁸ before collecting. Poly(A) RNA was purified through the use of oligo dT cellulose²¹. After electrophoresis, the RNA was blotted to NYTRAN Plus and probed with randomly primed ³²P-labelled PCR products. Hybridization signals were quantified with a Molecular Dynamics Phosphorimager.

Chromatin immunoprecipitation. YDS3 (wild-type), SRY3D2 (*hda1*), and SRYR35 (*rp3*) cells were grown to $A_{600} \sim 1.5$ in YEPD medium and chromatin was immunoprecipitated²² using lysis buffer containing 50 mM HEPES/KOH pH 7.5, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100%, 0.1% sodium deoxycholate, 0.1% SDS, and complete protease inhibitor (Boehringer). Immunoprecipitations were performed through the addition of 0.5 µl crude serum specific for each H4 acetylated lysine to 100 µl aliquots of extract. The antibodies used were either R728.5 (H4 acetyl-lysine 5), R12/8 (H4 acetyl-lysine 8), R101/12 (H4 acetyl-lysine 12) or R13/16 (H4 acetyl-lysine 16)²³. After the final wash with TE buffer, DNA was eluted and crosslinking was reversed by adding 100 µl elution buffer (50 mM Tris/HCl pH 8.0 and 10 mM EDTA/1% SDS) and incubating overnight at 65 °C. 50 ng purified DNA (as determined by fluorometry) was analysed by PCR²². Input DNA obtained from 20 µl extract was purified in parallel with the immunoprecipitated samples. PCR products were resolved by PAGE on 6% acrylamide::bis-acrylamide (19::1) TBE gels and visualized with ethidium bromide. For quantification of PCR products, ethidium-bromide-stained PCR products were photographed on Polaroid Type 55 film, scanned and quantified using NIH Image 1.58 software.

SIN3 disruption. A *sin3* mutation was generated in YDS3 (ref. 20) cells using the *sin3::URA3* disruption plasmid pMV117 (ref. 24) to produce strain NSY100. The disruption was confirmed by Southern blot analysis.

Received 19 January; accepted 26 March 1998.

- Hassig, C. A. *et al.* Histone deacetylase activity is required for full transcriptional repression by mSin3A. *Cell* **89**, 341–347 (1997).
- Laherty, C. D. *et al.* Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. *Cell* **89**, 349–356 (1997).

3. Kadosh, D. & Struhl, K. Repression by Ume6 involves recruitment of a complex containing Sin3 corepressor and Rpd3 histone deacetylase to target promoters. *Cell* **89**, 365–371 (1997).
4. Nagy, L. *et al.* Nuclear receptor repression mediated by a complex containing SMRT, mSin3A, and histone deacetylase *Cell* **89**, 373–380 (1997).
5. Heinzel, T. *et al.* A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. *Nature* **387**, 43–48 (1997).
6. Alland, L. *et al.* Role for N-CoR and histone deacetylase in Sin3-mediated transcriptional repression. *Nature* **387**, 49–55 (1997).
7. Rundlett, S. E. *et al.* HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription. *Proc. Natl Acad. Sci. USA* **93**, 14503–14508 (1996).
8. Slekar, K. H. & Henry, S. A. SIN3 works through two different promoter elements to regulate INO1 gene expression in yeast. *Nucleic Acids Res.* **23**, 1964–1969 (1995).
9. Strich, R. *et al.* UME5 is a key regulator of nitrogen repression and meiotic development. *Genes Dev.* **8**, 796–810 (1994).
10. Hecht, A., Strahl-Bolsinger, S. & Grunstein, M. Spreading of transcriptional repressor SIR3 from telomeric heterochromatin. *Nature* **383**, 92–96 (1996).
11. Kasten, M. M., Dorland, S. & Stillman, D. J. A large protein complex containing the yeast Sin3p and Rpd3p transcriptional regulators. *Mol. Cell. Biol.* **17**, 4852–4858 (1997).
12. Zhang, Y. *et al.* Histone deacetylases and Sap18, a novel polypeptide, are components of a human Sin3 complex. *Cell* **89**, 357–364 (1997).
13. Mitchell, A. P. Control of meiotic gene expression in *Saccharomyces cerevisiae*. *Microbiol. Rev.* **58**, 56–70 (1994).
14. Grunstein, M. Histone function in transcription. *Annu. Rev. Cell Biol.* **6**, 643–678 (1990).
15. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
16. Dean-Johnson, M. & Henry, S. A. Biosynthesis of inositol in yeast. Primary structure of myo-inositol-1-phosphate synthase (EC5.5.1.4) and functional analysis of its structural gene, the INO1 locus. *J. Biol. Chem.* **264**, 1274–1283 (1989).
17. Mann, R. K. & Grunstein, M. Histone H3 N-terminal mutations allow hyperactivation of the yeast GAL1 gene *in vivo*. *EMBO J.* **11**, 3297–3306 (1992).
18. Guthrie, C. & Fink, G. R. *Guide to Yeast Genetics and Molecular Biology* (Academic, San Diego, CA, 1991).
19. Kayne, P. S. *et al.* Extremely conserved histone H4 N terminus is dispensable for growth but essential for repressing the silent mating loci in yeast. *Cell* **55**, 27–39 (1988).
20. Laman, H., Balderes, D. & Shore, D. Disturbance of normal cell cycle progression enhances the establishment of transcriptional silencing in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **15**, 3608–3617 (1995).
21. Flomerfelt, F. A., Briehl, M. M., Dowd, D. R., Dieken, E. S. & Miesfeld, R. L. Elevated glutathione S-transferase gene expression is an early event during steroid-induced lymphocyte apoptosis. *J. Cell. Physiol.* **154**, 573–581 (1993).
22. Strahl-Bolsinger, S., Hecht, A., Luo, K. & Grunstein, M. SIR2 and SIR4 interactions differ in core and extended telomeric heterochromatin in yeast. *Genes Dev.* **11**, 83–93 (1997).
23. Clarke, D. J., O'Neill, L. P. & Turner, B. Selective use of H4 acetylation sites in the yeast *Saccharomyces cerevisiae*. *Biochem. J.* **294**, 557–561 (1993).
24. Vidal, M., Strich, R., Esposito, R. E. & Gaber, R. F. RPD1 (SIN3/UME4) is required for maximal activation and repression of diverse yeast genes. *Mol. Cell. Biol.* **11**, 6306–6316 (1991).

Acknowledgements. We thank S. Henry for discussions and R. Gaber for plasmid pMV117. This work was supported by Public Health Service grants from the NIH (to M.G.).

Correspondence and requests for materials should be addressed to M.G. (e-mail: mg@mbi.ucla.edu).

Structure of the calcium pump from sarcoplasmic reticulum at 8-Å resolution

Peijun Zhang*, Chikashi Toyoshima†, Koji Yonekura†‡, N. Michael Green§ & David L. Stokes*

* Skirball Institute of Biomolecular Medicine, New York University Medical Center, 540 First Avenue, New York, New York 10016, USA

† Institute of Molecular and Cellular Biosciences, The University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

§ National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

The calcium pump from sarcoplasmic reticulum (Ca^{2+} -ATPase) is typical of the large family of P-type cation pumps. These couple ATP hydrolysis with cation transport, generating cation gradients across membranes. Ca^{2+} -ATPase specifically maintains the low cytoplasmic calcium concentration of resting muscle by pumping calcium into the sarcoplasmic reticulum; subsequent release is used to initiate contraction. No high-resolution structure of a P-type pump has yet been determined, although a 14-Å structure of Ca^{2+} -ATPase, obtained by electron microscopy of frozen-hydrated, tubular crystals¹, showed a large cytoplasmic head

connected to the transmembrane domain by a narrow stalk. We have now improved the resolution to 8 Å and can discern ten transmembrane α -helices, four of which continue into the stalk. On the basis of constraints from transmembrane topology, site-directed mutagenesis and disulphide crosslinking, we have made tentative assignments for these α -helices within the amino-acid sequence. A distinct cavity leads to the putative calcium-binding site, providing a plausible path for calcium release to the lumen of the sarcoplasmic reticulum.

As in previous studies^{2–4}, tubular crystals were induced in rabbit sarcoplasmic reticulum by using decavanadate. We also included an inhibitor, dansyl thapsigargin⁵, that promotes crystallization by

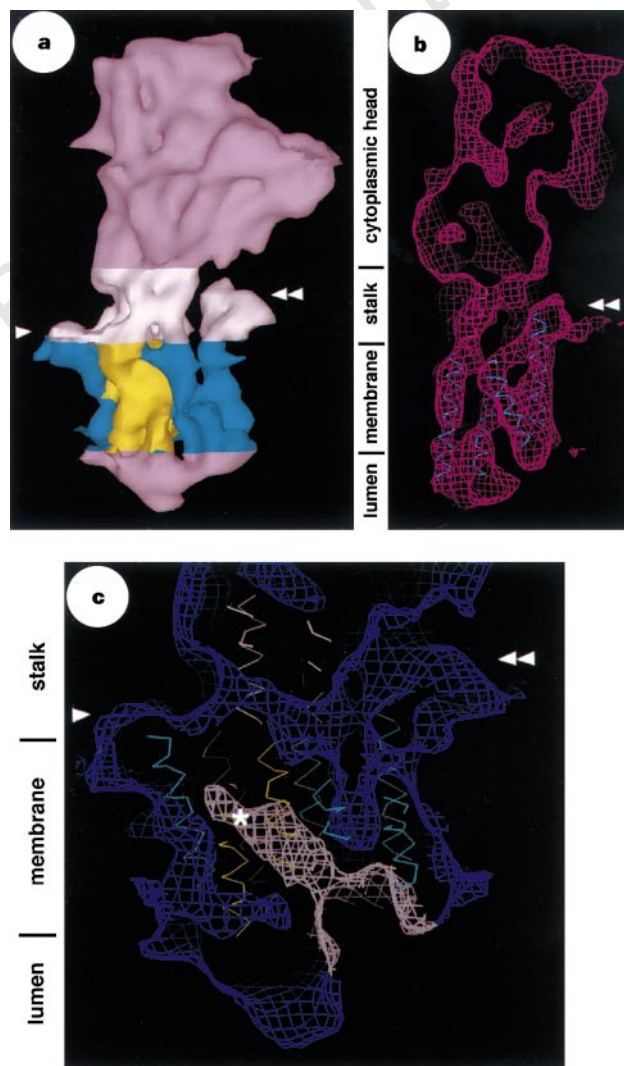


Figure 1 Ca^{2+} -ATPase density map at 8-Å resolution. **a**, Surface representation of the whole molecule with the raised platform (double arrowhead) and flat platform (single arrowhead) indicated. The cyan and yellow densities in the transmembrane domain correspond to the α -helices in **b**, **c** and Fig. 3. **b**, View of helices G, H, I and J (blue) fitted to the map; the 'tenth', highly inclined helix (I) is second from the left, and the helical hairpin composed by G and H is on the right. This view is rotated 90° relative to that in **a**, looking from the right at the model in **a**. **c**, A cavity within the transmembrane domain has been highlighted in grey and leads to putative calcium sites (asterisk). This contour level accounts for 100% of the molecular volume (147,000 Å³ based on a relative molecular mass of 110,000), whereas contour levels in **a** and **b** correspond to 75% of this volume. The view is the same as that in **a**, but only a section from the centre of the molecule is shown, to reveal the cavity. The lines composing the nets in **b** and **c** are at 2-Å intervals.

‡ Present address: Protonic Nanomachine Project, ERATO, JST, 3-4 Hikaridai, Seika, 619-02 Kyoto, Japan.

locking Ca^{2+} -ATPase in the so-called E_2 conformation, which is characterized by low-affinity calcium binding from the luminal side of the membrane⁶. Tubes were imaged in the frozen-hydrated state, and optical diffraction of the best images revealed a strong layer line at 10.5-Å resolution. The narrowest tubes (600 Å in diameter) were selected for image processing, which involved a combination of Fourier space averaging and real-space averaging (see Methods). The phase residuals indicated that the resulting data extended to at least 8-Å resolution (Table 1).

The resulting map (Figs 1, 2) reveals many rod-like densities, separated by ~10 Å, that presumably correspond to α -helices. Because the transmembrane and stalk domains are predicted to be entirely α -helical (Fig. 3), we have concentrated exclusively on these regions. Initially, α -carbon backbones were easily fitted to nine rod-like densities within the transmembrane domain (Fig. 2c–f), but this left a strong, unmatched density on the luminal side of the membrane (arbitrarily labelled 'I' in Fig. 2f). Given the strong consensus for ten transmembrane helices^{7–9}, a tenth, highly inclined

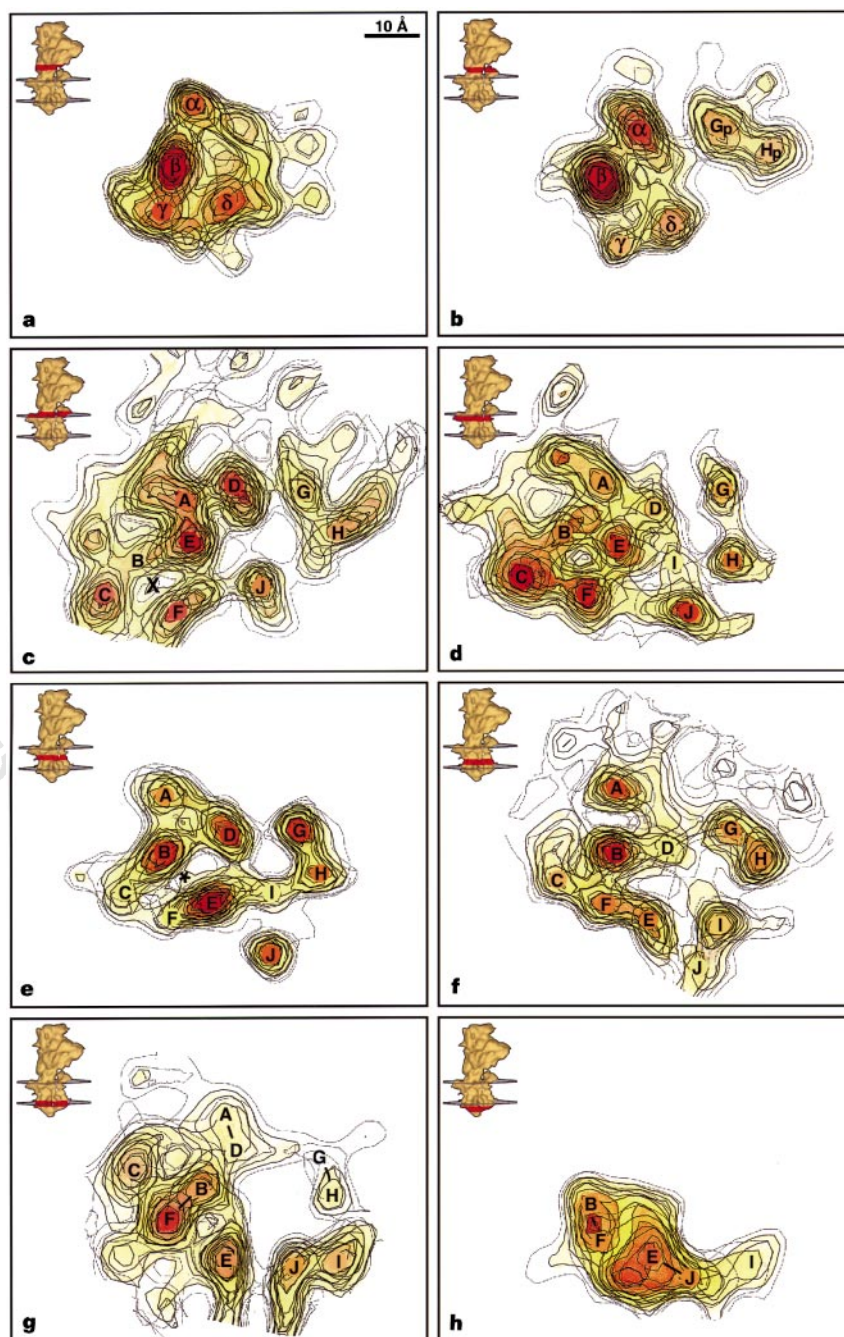


Figure 2 Density cross-sections through the transmembrane and stalk of Ca^{2+} -ATPase. Four sections (2 Å thick) were superimposed in each panel (a–h) and their position is shown on the inset surface model. The top of the cavity is marked by an asterisk in **e** at the same location as in Fig. 1c. The narrow passageway from

the cytoplasm is marked with a cross in **c**. The lowest, dotted contour corresponds to 100% volume recovery; yellow, orange and red indicate increasing density.

Table 1 Statistics of data averaging

	$n_{1,0}, n_{0,1}^*$			Real-space average	Edited average†
	-21,6	-22,6	-23,6		
No. of tubes	6	10	7	23	23
No. of molecules	16,368	22,976	12,400	51,744	51,744
Unit cell dimension					
<i>a</i> (Å)	57.1 ± 0.6	57.6 ± 0.2	57.9 ± 0.5		
<i>b</i> (Å)	116.6 ± 0.6	116.4 ± 1.7	115.8 ± 1.2		
γ (°)	63.9 ± 0.6	64.8 ± 0.5	65.2 ± 0.5		
Twofold phase residual (°)‡					
20.0 Å	7.4	5.9	8.7	3.1	3.1 (96)
14.4 Å	34.9	27.6	33.4	18.0	17.5 (91)
10.0 Å	42.3	38.9	43.1	31.9	29.6 (79)
8.0 Å	44.9	44.0	45.1	42.2	34.6 (57)

* Selection rule of the three independent data sets is characterized by the Bessel order (*n*) of the (1,0) and the (0,1) layer lines as defined in ref. 1.

† This data set, used for the final map, included 183 layer lines and was truncated at 8-Å resolution. It was derived from the real-space average by limiting the radial extent of each layer line to exclude noisy data with poor twofold phase residuals. Numbers in parentheses indicate the percentage of data remaining after editing. Overall phase residual was 16.2°.

‡ Amplitude-weighted phase residuals for twofold symmetry excluded equatorial data as well as data lower than 0.1% of the maximal off-equatorial amplitude (~99% of off-equatorial data were included). Random phases produce a phase residual of 45°.

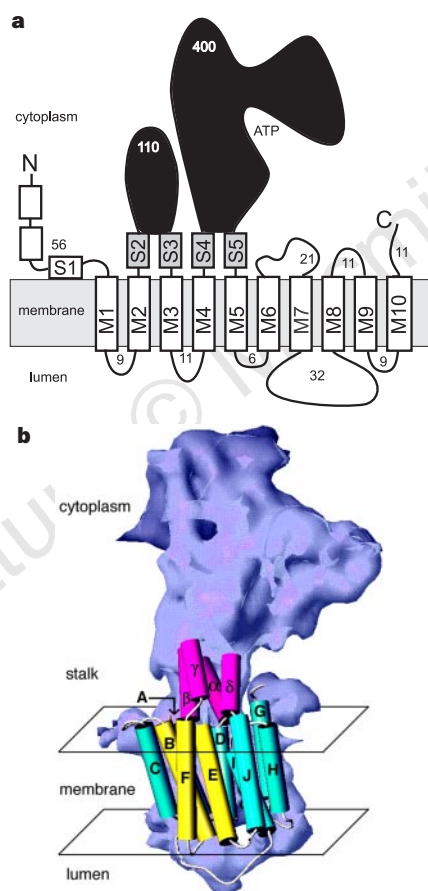


Figure 3 Transmembrane architecture of Ca²⁺-ATPase. **a**, Rectangles correspond to predicted α -helices and numbers indicate the number of residues in intervening loops. Stalk helices are grey and exclude S1, given that only four stalk segments are observed in our map. The two long black loops compose the cytoplasmic domains, which carry out ATP hydrolysis, but were not fitted. **b**, The result of fitting transmembrane and stalk helices to our density map. An important feature is the three-stranded coiled coil formed by the yellow helices, which give rise to the cavity seen in Fig. 1c as helices E and I diverge toward the luminal surface. All fitted helices were straight, although at higher resolution some might turn out to be bent. The white loops give a sense of continuity between helices based on the sequence assignments in Fig. 5a, but are not of accurate length.

helix was fitted to this last density. Unlike the initial nine helices, helix I does not traverse the membrane fully but meets helix E 6–8 Å from the cytoplasmic surface of the membrane (Fig. 2c, d). In general, only two of the ten helices (D and F) are inclined by <10°, five helices by 18° to 22°, and helices E and I by 27° and 34°, respectively, causing the relative positions of helices to change dramatically in sections parallel to the membrane plane (Fig. 2). Nevertheless, there are several clear groupings and a general trend toward right-handed twisting in these groups. In particular, pairwise associations occur between helices G and H, as well as between I and J, with the latter pair twisting around one another with a right-handed sense. A right-handed twist is seen again in the looser association between helices A and D and in the three-way association between helices B, E and F; at the cytoplasmic surface of the membrane, helix C is also associated with the latter group, but is excluded towards the luminal membrane surface. Between this threesome is a cavity (asterisk in Figs 1c, 2e) leading from the middle of the membrane to the luminal surface that is visible even at a contour level that includes 100% of the expected molecular volume (Fig. 1c) and appears to be surrounded exclusively by protein as it traverses the hydrophobic core of the bilayer. The presence of such a cavity is consistent with a water-filled channel that could provide access from the calcium-binding sites to the lumen, as would be expected for the E₂ conformation believed to populate this crystal form⁴. Such a channel has previously been hypothesized to explain voltage effects on sodium binding to the E₂ conformation of the related Na⁺/K⁺-ATPase^{10,11}, which suggest that cations move more than halfway across the membrane to reach their binding sites. Cytoplasmic access to the calcium sites should be restricted in the E₂ conformation, but even so there is a potential passageway starting at the cytoplasmic surface between helices B, C, E and F (cross in Fig. 2c) and leading down to the beginning of the larger cavity (asterisk in Fig. 2e). This passage is visible only at a higher density cutoff, which is consistent with the restricted access expected for cytoplasmic calcium in the E₂ conformation.

The stalk is 24 Å long and divided into four, rod-like densities, labelled α , β , γ and δ (Fig. 2b). The connection between these stalk densities and the transmembrane helices is obscured by the low contrast around the phosphate headgroups of the lipid bilayer (Fig. 2c). Nevertheless, the trajectory of α , β and δ densities suggest that they connect with transmembrane helices D, E and J, respectively; β and δ are almost collinear with E and J, whereas α is inclined at ~40° relative to D. According to structure predictions⁷ (Fig. 3a), the stalk is composed of four or five α -helices, and we found that α -helical backbones fit plausibly into α , β and δ densities. However, the fourth stalk density, γ , is smaller, has lower density throughout the stalk and a tenuous connection with transmembrane helix F; thus the identification of γ as a stalk helix and its association with F remains tentative. Densities G_p and H_p (Fig. 2b) form a raised platform on the cytoplasmic surface of the membrane directly above helices G and H; as seen in Fig. 1 (double arrowhead), these densities are set apart from the stalk and do not connect with the main cytoplasmic head, and so are less likely to connect transmembrane sequences with either of the two, large cytoplasmic loops (Fig. 3). A second, flat platform caps transmembrane helices A, B and C at the cytoplasmic membrane surface to the left of the stalk (single arrowhead in Fig. 1). On the luminal side of the membrane, a small, compact domain bridges helices I and J to a larger group containing helices B, C, E and F (Figs 1a and 2g, h). Helices G and H come together at the luminal surface and have only weak associations with surrounding densities, and so resemble a helical hairpin (Fig. 1b).

We then tried to associate transmembrane and stalk densities with helices predicted from the amino-acid sequence (Fig. 3). These predictions⁷ are supported by the current evidence for transmembrane topology^{8,12} and provide two important constraints in the form of direct connections between stalk helices and particular

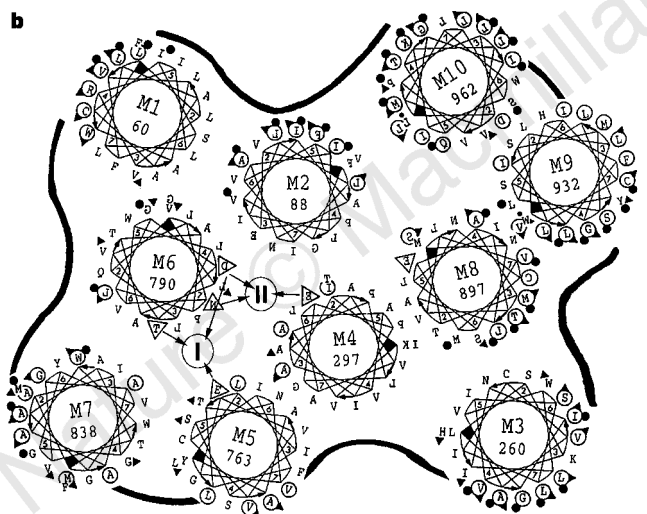
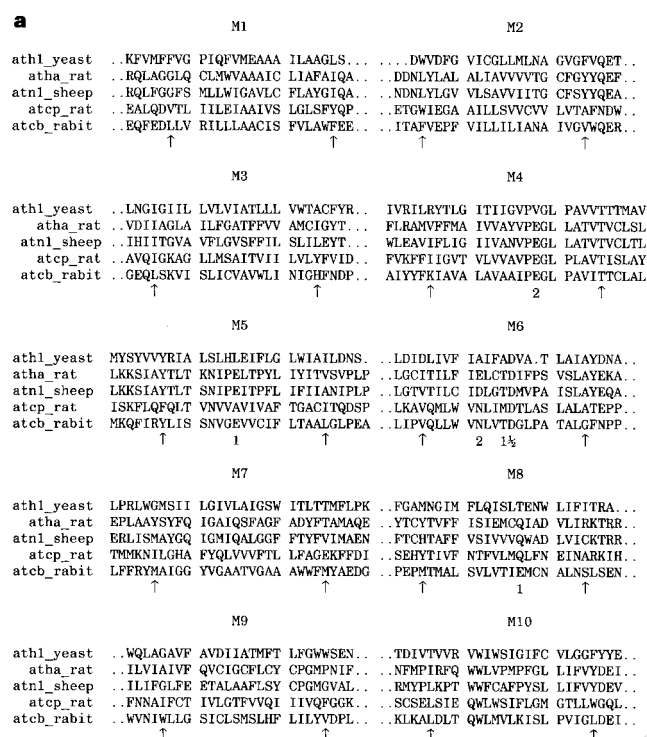
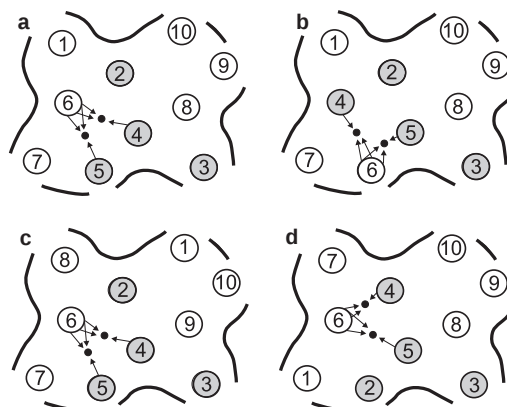


Figure 4 Variability within the transmembrane sequences of P-type ion pumps. **a**, Sequence alignment between representatives of several subfamilies of P-type pumps: from the top, yeast H⁺-ATPase, rat H⁺/K⁺-ATPase, sheep Na⁺/K⁺-ATPase, rat plasma-membrane Ca²⁺-ATPase and rabbit sarcoplasmic-reticulum Ca²⁺-ATPase. These alignments were done according to refs 19, 25, but M8 for the PMCA and Na⁺/K⁺ families were realigned to match Q with E 908 from SERCA1²⁶. Arrows define residues plotted in **b**; 1 and 2 indicate residues that contribute to calcium sites I and II⁹, respectively. **b**, Residue conservation was analysed within three subfamilies¹⁹ and variable residues are indicated by encircled letters for SERCA pumps, by filled circles for PMCA pumps, and by filled triangles for the Na⁺/K⁺-ATPase pumps. The helical wheels were superimposed on the density map (Fig. 2e) according to our most favoured assignment (Fig. 5a), and are oriented to expose variable faces to the lipid¹⁸ and to juxtapose calcium ligands¹³ (letters in triangles) and cysteine crosslinks¹⁷ in M4, M5 and M6. Letters correspond to the sequence of SERCA1; helices with normal letters are running away from, and those with mirrored letters towards the cytoplasm. The first residue in each helix is numbered at the centre and marked with a filled diamond; the first heptad along the helix is numbered (2–7) and linked to corresponding residues in successive heptads by the curved arrows.

transmembrane helices, and close proximity between several pairs of transmembrane helices connected by short loops. Further constraints come from the results of site-directed mutagenesis, which localized the calcium-binding sites between M4, M5 and M6 (refs 9, 13–16), and from disulphide links between pairs of cysteines introduced into M4 and M6 (ref. 17). The positions of the linked cysteines support a right-handed coiling of M4 and M6 over several helical turns. Residue variability can also shed light on helix packing, given the tendency for variable sites to face the lipid and for conserved sites to participate in helix–helix interactions¹⁸. We have extended a previous analysis¹⁹ by aligning transmembrane segments from several families of P-type pumps (Fig. 4a), using a few well-conserved markers within, or just outside, the hydrophobic segments and assuming an absence of gaps. The variable sites (Fig. 4b) of Na⁺/K⁺-ATPases (triangles) and plasma-membrane Ca²⁺-ATPases (circles) correspond well with those of SR Ca²⁺-ATPases (encircled letters), supporting the existence of a common transmembrane structure for the three families, even though there is less than 20% identity between the sequences of the transmembrane segments. In terms of helix packing, M4 and M6 are clearly the most highly conserved helices and should therefore have minimal exposure to bulk lipid. M5 is slightly less conserved, followed by M1, M2 and M8. M3, M9 and M10 are highly variable and the variable residues fall on one side of these helices, defining a face likely to be exposed to lipid. The specific positions of helices in Fig. 4b correspond to a section through the middle of the membrane at



the level expected for calcium-binding sites ($\sim 12 \text{ \AA}$ from the cytoplasmic surface; Fig. 2e); although the relative positions of helices change through the membrane, the exposed faces are generally preserved (Fig. 2).

Unfortunately, these constraints and considerations are not sufficient to assign unambiguously the sequence of transmembrane helices in our map. Nevertheless, we favour one set of assignments (Figs. 4b and 5a) over several others (Fig. 5b–d), which together provide specific hypotheses to be tested by further mutagenesis and crosslinking studies. Our preference is based on the most likely connections between stalk helices and transmembrane helices, for example $\alpha \rightarrow D$, $\beta \rightarrow E$, $\gamma \rightarrow F$ and $\delta \rightarrow J$ (Fig. 3b). The connection between γ and F is the most tenuous and, if correct, would probably involve an unstructured region near the membrane surface. Transmembrane density E is most central and so has been associated with the most highly conserved, stalk-associated helix, M4. Densities B, E and F form a right-handed coiled-coil that surrounds the distinctive cavity, and have thus been assigned to M4, M5 and M6. The geometry of this arrangement (Fig. 4b) is consistent with the mutagenesis results that associate particular residues with either of the two calcium-binding sites and specifically place them side by side at the intersection of channels from the cytoplasmic and luminal sides of the membrane⁹. Assignment of M8 to the highly inclined density I allows E908 to project into the latter channel, consistent with its effects on calcium transport^{9,13,14}; M8 would thus be rather short, perhaps owing to the unwinding of its polar carboxy terminus (NSLSEN; Fig. 4a). Density J is very exposed to lipid yet clearly extends into the stalk, making it a good candidate for the variable M3 helix. The extreme variability of M9 and M10 make this helical hairpin an attractive candidate for the isolated pair of transmembrane densities G and H, which are clearly associated at the luminal side of the membrane (Fig. 1b). Indeed, these assignments are consistent with the luminal disposition of several transmembrane densities (Fig. 2f–h), although the visualization of short loops at this resolution may in itself be unreliable. The loop between M7 and M8 is the longest, composing the bulk of the luminal domain and presenting an exposed, antigenic site^{20,21}, consistent with M7 and M8 being on opposite sides of the molecule. An apparent discrepancy arises from putting the larger amino terminus in the flat platform and the smaller C terminus in the raised platform. This could be explained either by a disordered N terminus, or by putting the N terminus in the raised platform and hypothesizing a disordered or unstructured connection to M1; in the latter case, the arch of density connecting G with A in Fig. 2c could represent this connection.

Thus, we have resolved ten helices within the transmembrane domain of Ca^{2+} -ATPase and tentatively identified the site of calcium binding. Consistent with the conformation in this crystal form, a large cavity provides access to this site from the lumen, with a constricted passageway leading to the cytoplasmic surface. This suggests an oblique overall path for calcium through the molecule, and structural studies of the alternative E_1 conformation will be required to define the structural changes that drive transport along this path. □

Methods

Crystals were prepared as described¹, except that $30 \mu\text{M}$ dansyl thapsigargin⁵ was added during crystallization. New strategies were used for processing images. To begin, a reference data set was compiled from the two best images in each of three symmetry groups ($n = 6$ for (0,1) layer line and $n = -21, -22$ or -23 for (1,0) layer lines according to previous nomenclature¹) using established methods. Individual tubes were divided into short stretches ($\sim 1,000 \text{ \AA}$) and positional parameters refined by comparing Fourier data with data from the corresponding reference data set²². This procedure corrected slight stretching and bending of the tubes, led to improved phase statistics over previous methods, and allowed the use of longer areas along slightly bent tubes, improving signal-to-noise ratios significantly. Initial defocus levels were

determined from Fourier amplitudes obtained either from nearby carbon film or from the tube itself²³, and were later refined by comparing phases from the individual tubes with those in the averaged data sets. Data from each symmetry group were averaged and compensated for the contrast transfer function assuming 4.6% amplitude contrast²⁴. These three independent averaged data sets were edited to exclude noisy data, and were truncated at either 14- or 10- $\text{\AA}$ resolution before calculating three-dimensional maps. After adjusting their relative magnifications, molecules within the three unit cells were masked and aligned in real space by cross-correlation; at this stage, twofold symmetry was not enforced and alignment parameters were determined independently for the twofold related molecules composing the unit cell. New maps were then created from a full set of unedited data to 6.5- $\text{\AA}$ resolution, aligned according to these same alignment parameters, weighted according to the square root of the number of contributing molecules, and averaged in real space (Table 1). This averaged map was then used to generate a set of Fourier data with the helical symmetry of the alignment reference ($n = -22$ for (1,0) layer line), which were edited and truncated to 8- $\text{\AA}$ resolution before calculating the final map.

Received 3 December 1997; accepted 20 January 1998.

- Toyoshima, C., Sasabe, H. & Stokes, D. L. Three-dimensional cryo-electron microscopy of the calcium ion pump in the sarcoplasmic reticulum membrane. *Nature* **362**, 469–471 (1993).
- Dux, L. & Martonosi, A. Two-dimensional arrays of proteins in sarcoplasmic reticulum and purified Ca^{2+} -ATPase vesicles treated with vanadate. *J. Biol. Chem.* **258**, 2599–2603 (1983).
- Taylor, K. A., Dux, L. & Martonosi, A. Three-dimensional reconstruction of negatively stained crystals of the Ca^{2+} -ATPase from muscle sarcoplasmic reticulum. *J. Mol. Biol.* **187**, 417–427 (1986).
- Stokes, D. L. & Lacapere, J.-J. Conformation of Ca^{2+} -ATPase in two crystal forms: Effects of Ca^{2+} , thapsigargin, AMP-PCP, and Cr-ATP on crystallization. *J. Biol. Chem.* **269**, 11606–11613 (1994).
- Hua, S., Malak, H., Lakowicz, J. R. & Inesi, G. Synthesis and interaction of fluorescent thapsigargin derivatives with the sarcoplasmic reticulum ATPase membrane-bound region. *Biochemistry* **34**, 5137–5142 (1995).
- Sagara, Y., Wade, J. B. & Inesi, G. A conformational mechanism for formation of a dead-end complex by the sarcoplasmic reticulum ATPase with thapsigargin. *J. Biol. Chem.* **267**, 1286–1292 (1992).
- MacLennan, D. H., Brandl, C. J., Korczak, B. & Green, N. M. Amino-acid sequence of a Ca^{2+} + Mg^{2+} -dependent ATPase from rabbit muscle sarcoplasmic reticulum, deduced from its complementary DNA sequence. *Nature* **316**, 696–700 (1985).
- Moller, J. V., Juul, B. & le Maire, M. Structural organization, ion transport, and energy transduction of ATPases. *Biochim. Biophys. Acta* **1286**, 1–51 (1996).
- MacLennan, D. H., Rice, W. J. & Green, N. M. The mechanism of Ca^{2+} transport by sarco(endo)-plasmic reticulum Ca^{2+} ATPases. *J. Biol. Chem.* **272**, 28815–28818 (1997).
- Gadsby, D. C., Rakowski, R. F. & De Weer, P. Extracellular access to the Na, K pump: Pathway similar to ion channel. *Science* **260**, 100–103 (1993).
- Hilgemann, D. W. Channel-like function of the Na,K pump probed at microsecond resolution in giant membrane patches. *Science* **263**, 1429–1432 (1994).
- Stokes, D. L., Taylor, W. R. & Green, N. M. Structure, transmembrane topology and helix packing of P-type ion pumps. *FEBS Lett.* **346**, 32–38 (1994).
- Clarke, D. M., Loo, T. W., Inesi, G. & MacLennan, D. H. Location of high affinity Ca^{2+} -binding sites within the predicted transmembrane domain of the sarcoplasmic reticulum Ca^{2+} -ATPase. *Nature* **339**, 476–478 (1989).
- Andersen, J. P. Dissection of the functional domains of the sarcoplasmic reticulum Ca^{2+} -ATPase by site-directed mutagenesis. *Biochim. Biophys. Acta* **15**, 243–261 (1995).
- Rice, W. J. & MacLennan, D. H. Scanning mutagenesis reveals a similar pattern of mutation sensitivity in transmembrane sequences M4, M5, and M6, but not in M8, of the Ca^{2+} -ATPase of sarcoplasmic reticulum (SERCA1a). *J. Biol. Chem.* **271**, 31412–31419 (1996).
- Chen, L. *et al.* Short and long range functions of amino acids in the transmembrane region of the sarcoplasmic reticulum ATPase: a mutational study. *J. Biol. Chem.* **271**, 10745–10752 (1996).
- Rice, W. J. & MacLennan, D. H. Site-directed disulfide mapping of helices M4 and M6 in the Ca^{2+} binding domain of SERCA1a, the Ca^{2+} ATPase of fast-twitch skeletal muscle sarcoplasmic reticulum. *J. Biol. Chem.* **272**, 31412–31419 (1997).
- Baldwin, J. M. The probable arrangement of the helices in G protein-coupled receptors. *EMBO J.* **12**, 1693–1703 (1993).
- Green, N. M. In *The Sodium Pump: Structure, Mechanisms, Hormonal Control and its Role in Disease* (eds Bamberg, E. & Schoner, W.) 110–119 (Steinkopff, Darmstadt, 1994).
- Matthews, I., Sharma, R. P., Lee, A. G. & East, J. M. Transmembrane organization of (Ca^{2+} - Mg^{2+})-ATPase from sarcoplasmic reticulum: evidence for luminal location of residues 877–888. *J. Biol. Chem.* **265**, 18737–18740 (1990).
- Clarke, D. M., Loo, T. W. & MacLennan, D. H. The epitope for monoclonal antibody A20 (amino acids 870–890) is located on the luminal surface of the Ca^{2+} -ATPase of sarcoplasmic reticulum. *J. Biol. Chem.* **265**, 17405–17408 (1990).
- Beroukhim, R. & Unwin, N. Distortion correction of tubular crystals: Improvements in the acetylcholine receptor structure. *Ultramicroscopy* **70**, 57–81 (1997).
- Tani, K., Sasabe, H. & Toyoshima, C. A set of computer programs for determining defocus and astigmatism in electron images. *Ultramicroscopy* **65**, 31–44 (1997).
- Toyoshima, C., Yonekura, K. & Sasabe, H. Contrast transfer for frozen-hydrated specimens: II. Amplitude contrast at very low frequencies. *Ultramicroscopy* **48**, 165–176 (1993).
- Green, N. M. ATP-driven cation pumps: alignment of sequences. *Biochem. Soc. Trans.* **17**, 970–972 (1989).
- Guerini, D., Foletti, D., Vellani, F. & Carofoli, E. Mutation of conserved residues in transmembrane domains 4, 6, and 8 causes loss of Ca^{2+} transport by the plasma membrane Ca^{2+} pump. *Biochemistry* **35**, 3290–3296 (1996).

Acknowledgements. We thank R. Beroukhim and N. Unwin for the use of their programs for helical image analysis, and G. Inesi for the dansyl thapsigargin. This work was partly supported by the NIH and the Ministry of Education, Science, Sports and Culture of Japan.

Correspondence and requests for materials should be addressed to D.L.S. (e-mail: stokes@saturn.med.nyu.edu).

Three-dimensional map of the plasma membrane H⁺-ATPase in the open conformation

Manfred Auer^{*,†}, Gene A. Scarborough[‡] & Werner Kühlbrandt^{*}

^{*} Max-Planck-Institut für Biophysik, Abteilung Strukturbiologie, Heinrich-Hoffmann-Strasse 7, 60528 Frankfurt am Main, Germany

[†] European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

[‡] Department of Pharmacology, University of North Carolina, Chapel Hill, North Carolina 27599, USA

The H⁺-ATPase from the plasma membrane of *Neurospora crassa* is an integral membrane protein of relative molecular mass 100K, which belongs to the P-type ATPase family that includes the plasma membrane Na⁺/K⁺-ATPase and the sarcoplasmic reticulum Ca²⁺-ATPase. The H⁺-ATPase pumps protons across the cell's plasma membrane using ATP as an energy source, generating a membrane potential in excess of 200 mV (refs 1–3). Despite the importance of P-type ATPases in controlling membrane potential and intracellular ion concentrations, little is known about the molecular mechanism they use for ion transport. This is largely due to the difficulty in growing well ordered crystals and the resulting lack of detail in the three-dimensional structure of these large membrane proteins. We have now obtained a three-dimensional map of the H⁺-ATPase by electron crystallography of two-dimensional crystals grown directly on electron microscope

grids. At an in-plane resolution of 8 Å, this map reveals ten membrane-spanning α -helices in the membrane domain, and four major cytoplasmic domains in the open conformation of the enzyme without bound ligands.

Hydropathy analysis of the primary sequence of H⁺-ATPase suggested that there are 8–12 transmembrane spans^{1,4}; there is a large hydrophilic portion on the cytoplasmic side of the membrane which includes the amino and carboxy termini and two long loop segments^{5–9}. The exocytosomal part appears to consist mainly of short stretches that connect the transmembrane segments. The molecule is thus highly asymmetric, with a small extracellular portion, a transmembrane region containing about one quarter of the protein, and a cytoplasmic region that accounts for the largest part of the enzyme. The enzyme is fully functional in proton transport as a monomer¹⁰ but forms stable hexamers when isolated after detergent solubilization¹¹.

Images of two-dimensional crystals grown on the carbon-support film were recorded at tilt angles of 0° to 60° and processed, yielding amplitudes and phases of structure factors. A three-dimensional map was calculated from data extending to 8 Å resolution (Table 1 and Fig. 1). A cross-section of the map (Fig. 2a) indicates that the crystals consist of two layers of hexamers, separated by a narrow region of low density, and related to one another by two-fold symmetry. Within each layer, regions of tightly packed rod-shaped regions of high density, characteristic of transmembrane α -helices at 6 Å–8 Å resolution^{12–17}, define the membrane-spanning domain of the ATPase. The large cytoplasmic regions of the enzyme extend from the membrane domains towards the two-dimensional crystal surface.

A section parallel to the crystal plane at the level of the membrane reveals a striking pattern of circular units tightly packed on a hexagonal lattice, each representing one ATPase hexamer (Fig.

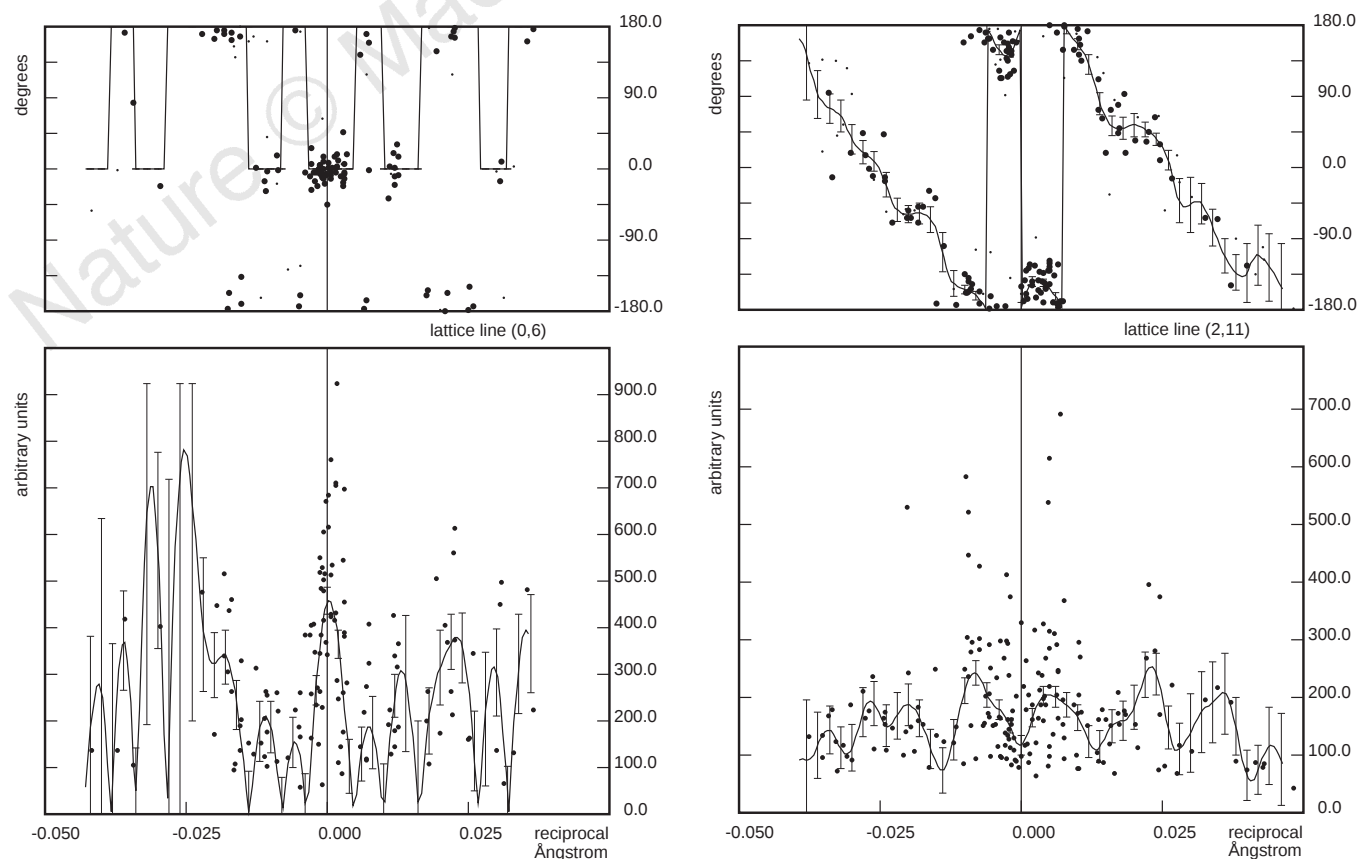


Figure 1 Phases (top panels) and amplitudes of two lattice lines. Individual data points show amplitudes and phases from processed images. Error bars indicate

the fitting error at the sampling points and are spaced 0.002 Å⁻¹ along the z*-axis, normal to the crystal plane.

Table 1 Crystallographic data

Symmetry	<i>P</i> 321
Lattice constants	$a = b = 167 \text{ \AA}$ $\gamma = 120^\circ$ $\sim 240 \text{ \AA}$
Thickness	
Number of images	60*
In-plane resolution	8 $\text{\AA}$
Maximum tilt angle	60.4°
Number of merged amplitudes and phases	30,313
Number of independent structure factors	3,561
Phase residual (90° is random)	26.2° (200–8 $\text{\AA}$) 22.2° (200–14 $\text{\AA}$) 32.1° (14–10 $\text{\AA}$) 37.7° (10–8.2 $\text{\AA}$) 40.6° (8.2–8.0 $\text{\AA}$)

* Distribution of tilt angles: 19 images (0°–10°), 9 (10°–20°), 13 (20°–30°), 11 (30°–45°), 8 (45°–60°).

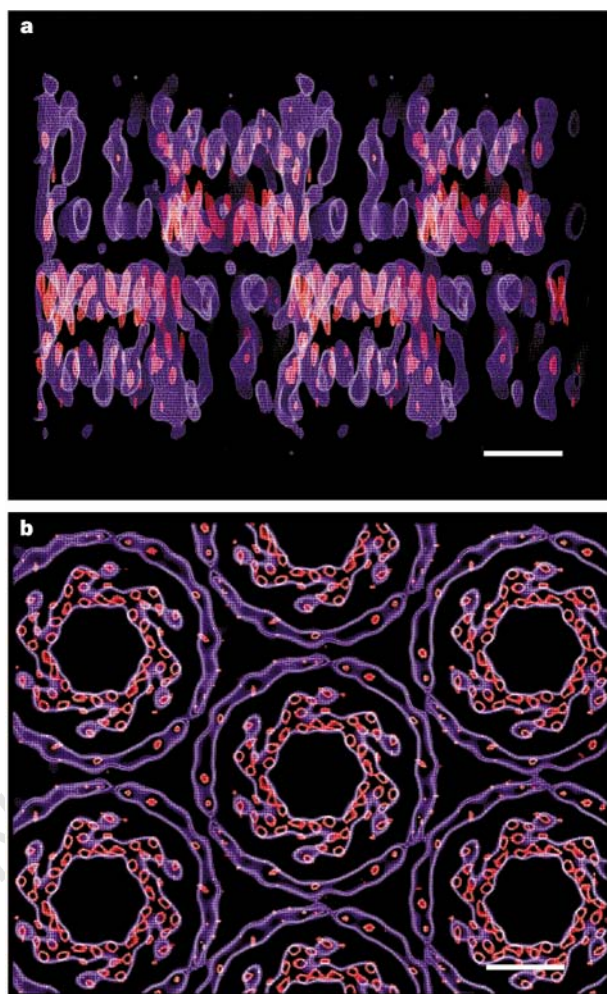


Figure 2 Three-dimensional map of the H^+ -ATPase. The red 8 $\text{\AA}$ map was contoured at 3 standard deviations to reveal high-density features, in particular membrane-spanning α -helices. The blue map was calculated from data truncated to 14 $\text{\AA}$ and contoured at 1.0 standard deviations to reveal the low-resolution envelope of the molecule. The 8 $\text{\AA}$ map shown here was sharpened by applying a temperature factor of $-500 \text{ \AA}^2$. **a**, Cross section through the 2-D crystal. The two layers of hexamers separated by a central, narrow region of low density are easily identified. Each hexamer contains a bundle of red transmembrane helices. **b**, Section parallel to the crystal plane through one layer of hexamers at the level of the membrane domain. Hexamers appear as two concentric rings of density. The inner ring contains the membrane domains of the six monomers, defined by the blue 14 $\text{\AA}$ outline. The red map indicates the position of transmembrane helices in cross-section within each monomer. The inner ring of helices is surrounded by a uniform ring of lower density which apparently contains the sugar head-groups of the dodecyl maltoside detergent. The detergent rings of each layer touch one another. Scale bar, 50 $\text{\AA}$.

2b). In this section, the hexamers appear as two concentric rings with outer diameters of 115 $\text{\AA}$ and 165 $\text{\AA}$, respectively. The inner ring has apparent 6-fold symmetry and consists of cross-sections of the rod-like α -helical densities seen in Fig. 2a, with the six repeating units representing the transmembrane domains of the six monomers. The density of the outer ring is lower and more uniform, and does not show a pronounced sixfold modulation. The distance between the outer edges of the outer and inner rings is $\sim 23 \text{ \AA}$ – $28 \text{ \AA}$, which is roughly the length of an extended dodecyl maltoside molecule. The outer ring therefore probably contains the sugar head-groups of the dodecyl maltoside detergent, which thus forms a single toroid micelle around each hexamer. Cross-sections of the detergent rings can be seen in Fig. 2a. The ATPase hexamers and the outer ring were already evident in a projection map obtained from crystalline single layers¹⁸.

Contacts between the hexamers in one layer exist at the level of the detergent ring (Fig. 2b) and between the cytoplasmic domains. Monomers within the hexamer interact through transmembrane

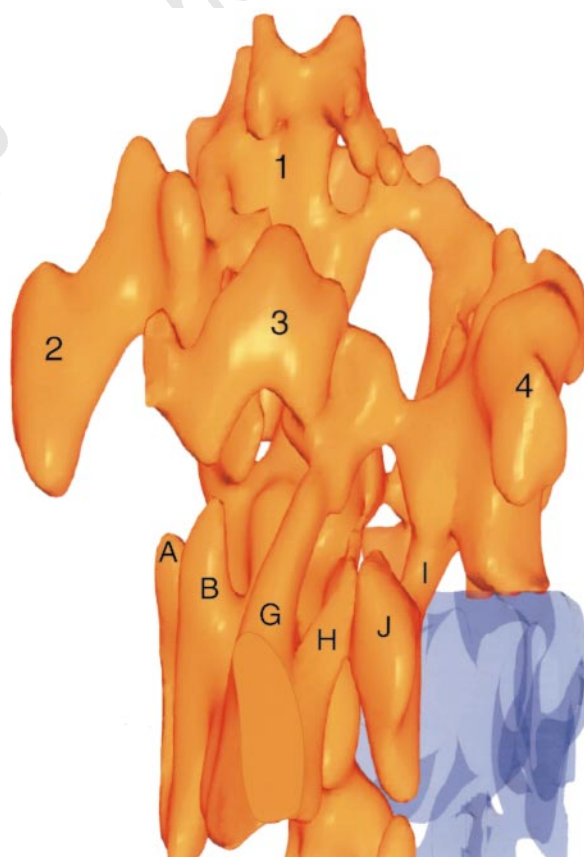


Figure 3 Side view of the H^+ -ATPase monomer, showing the average of two monomers related by non-crystallographic two-fold symmetry. The map was calculated without applying a temperature factor and contoured at 4 standard deviations, giving a volume of $\sim 64,000 \text{ \AA}^3$, corresponding to 50% of the relative molecular mass of a 100K protein^{29,30}. The transparent blue part shows the detergent ring. Subdomains in the large cytoplasmic region are labelled 1-4; transmembrane helices in the membrane domain, A-J.

helices at either end of the molecule which approach to within 10 Å (centre-to-centre), and through a small, elongated protrusion in the cytoplasmic region (see below). These interactions may explain the unusual stability of the H⁺-ATPase hexamers¹¹.

Close inspection of a superposition of two neighbouring monomers in the hexamer reveals that their main features are identical to within the sampling accuracy of the map (~2 Å). We conclude that these monomers are related by non-crystallographic two-fold symmetry, which together with the crystallographic 3-fold axis accounts for the 6-fold axis running centrally through the hexamer. An averaged map of both monomers is shown in Fig. 3, which indicates a total height of the molecule of ~120 Å, and a maximum width of ~85 Å. The exocytosolic portion is quite small, consistent with hydrophathy plots predicting the topography of the polypeptide^{1,4} and with biochemical data^{1,2,5-9}.

The membrane domain at the level of the detergent ring is defined by a compact bundle of 10 membrane-spanning α-helices which are

labelled A–J in Fig. 3 and 4. It covers an area of 30 Å × 40 Å and is 35 Å–40 Å thick, as judged by the average extent of the membrane-spanning helices in the z direction (Fig. 4b). The apparent length of individual helices varies considerably. Most of them are ~35 Å long, as expected. Helices H and J appear shorter, but this is probably due to a combination of weaker density and limited resolution in the z direction. Helices closest to the non-crystallographic 6-fold axis have particularly well defined densities which may reflect greater rigidity of this part of the hexamer.

The transmembrane helices are tightly packed, with centre-to-centre distances ranging from 8 Å–13 Å. None of them runs exactly perpendicular to the membrane plane (Fig. 4a). Tilt angles of helices B, C, D, F, G and J range from ~5°–15°, whereas helices A, E, H and I are more highly tilted by ~16°–22° (Fig. 4b). Helices A, B, C and E appear to form a 4-helix bundle with a right-handed twist (Fig. 4a). Helices D, F, H and G form a second layer on one side of the bundle, their orientation being similar to that of helices C and E. Helices I and J, which again have the same orientation, can then be thought of as forming a third layer. Most helices are essentially straight, except for helices E and I, which are slightly curved. On the exocytosolic surface, a loop connecting helices D and F extends ~6 Å beyond the outer membrane surface. This loop makes contact with the corresponding loop in a monomer in the other layer of the two-dimensional crystal (Fig. 2a).

The cytoplasmic region has an open and complex structure. It is continuous with the membrane domain from which it extends upwards to a maximum height of ~80 Å above the membrane surface, and measures ~85 Å × 50 Å in cross-section. It is divided into four distinct domains, labelled 1 to 4 in Fig. 3. Domain 1 contains the bulk of the cytoplasmic part, with a cross-section of ~35 Å × 25 Å. Domain 2 is a hook-shaped extension of domain 1, pointing towards the centre of the hexamer. It is attached to domain 3 of the next monomer, thus contributing to the tight hexamer contacts. Domain 3 is separated from domain 1 by a deep ~10 Å cleft and is shaped like a flattened arch. Domain 4 is separated from domains 1 and 2 by a distance of up to ~15 Å and has the most complex shape. Strikingly, it extends down to the detergent ring, which means that in the cell it would be close to and probably in contact with the cytoplasmic membrane surface.

The *Neurospora* H⁺-ATPase described here is in its unliganded state. This state is thought to have a relatively open conformation, as ~175 amino-acid residues on the enzyme surface become occluded on binding certain ligands¹⁹. This would be consistent with our three-dimensional map showing an open conformation of the enzyme, which is expected to reorganize into a compact structure upon binding substrates and transition-state analogues^{1,2}. Such ligand-induced conformational changes are probably important in the ion-occlusion and molecular mechanism of ion translocation^{1,2}. Comparing the structure of the H⁺-ATPase to the 14 Å map of the Ca²⁺-ATPase²⁰, we find that both have similar overall dimensions and that their transmembrane parts look similar. However, the cytoplasmic region of the Ca²⁺-ATPase has a compact wedge-shaped head, linked to the membrane by a stalk, whereas that of the H⁺-ATPase looks more open, with several apparent connections to the membrane. We conclude that the two forms of ATPase differ substantially in their cytoplasmic regions. The overall identity between the polypeptide sequences of the two enzymes is 19% (ref. 21), but sequences assigned to those major portions of the cytoplasmic region that are common to both ATPases have substantial homology (data not shown). The cytoplasmic portions should thus be even more similar than the membrane domains. The compact cytoplasmic region of the Ca²⁺-ATPase may therefore represent a different conformational state for this class of enzymes from that of the H⁺-ATPase presented here. □

Methods

Two-dimensional (2D) crystals were grown from a solution that yields

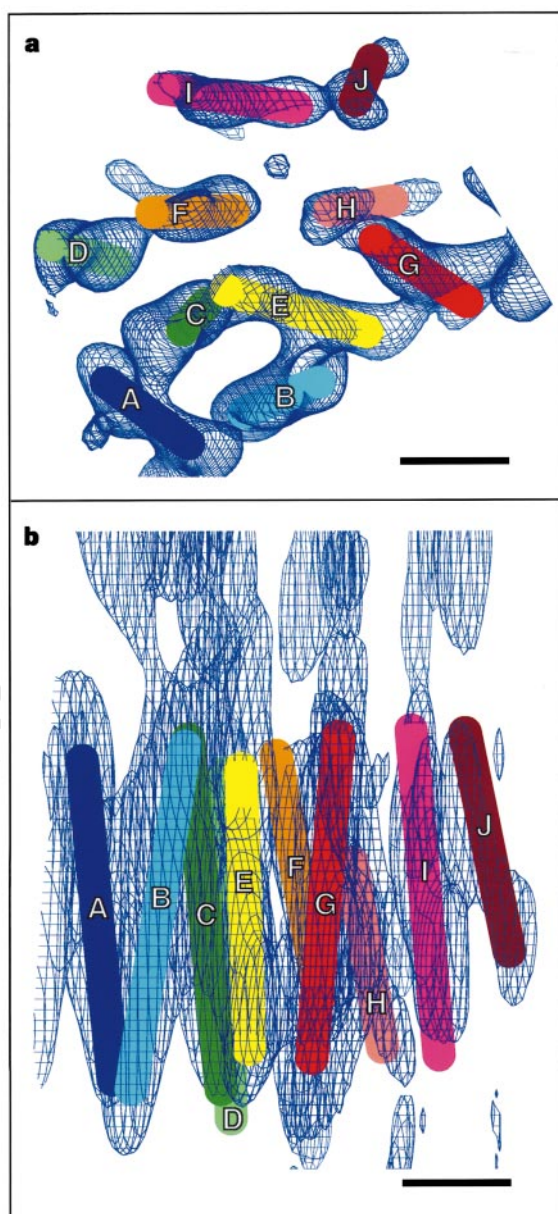


Figure 4 The transmembrane region. Top view (a) and side view (b) of the transmembrane region calculated on a 1 Å grid and sharpened by applying a temperature factor of ~500 Å². The map is contoured at 1.2 standard deviations. Helices A to J are indicated by coloured rods. Scale bar, 10 Å.

three-dimensional (3D) crystals²², containing H⁺-ATPase (1 mg ml⁻¹), dodecylmaltoside (1.28 mg ml⁻¹), threhalose (0.1%), ammonium sulphate (100 mM) and PEG4000 (Fluka) (10.5% w/v) in 30G solution¹⁸, on carbon-coated electron-microscope grids. Details of the crystallization procedure will be given elsewhere.

Grids were blotted for 5 to 10 s, frozen in liquid ethane²³, transferred at liquid-nitrogen temperature into a Gatan cryotransfer specimen holder and observed at ~100 K in a Philips CM200 FEG electron microscope operated at 200 kV acceleration voltage. Other grids were mounted in a modified Leica KF80 rapid-freezing apparatus for transfer into a JEOL 3000SFF electron microscope operated at 300 kV, equipped with a top-entry cryo-stage cooled to 4 K with liquid helium. Images were taken with 1 or 2 s exposure time. Most images recorded with the Philips CM200FEG were taken using a spot-scan procedure (I. Tews and W.K., unpublished) at a total dose of ~10 eÅ⁻². Images at tilt angles above 45° were recorded with the JEOL 3000 SFF at a total dose of ~25 eÅ⁻² in flood-beam mode. At these high tilt angles, the increased mean free path of 300 kV electrons was an advantage, as the effective specimen thickness at 60° tilt is ~500 Å. Images were recorded under low-dose conditions at a magnification of 50,000 on Kodak SO-163 film and developed for 12 min with full-strength Kodak D19.

Micrographs were screened by optical diffraction. 60 out of ~600 images were selected. Areas of 8,000 × 8,000 pixels were digitized using a Zeiss SCAI scanner with a 7-μm step size. Images were processed on a DEC/Alpha workstation using MRC image-processing programs²⁴. Phase comparison of symmetry-related reflections from untitled images indicated P321 symmetry. Only crystals with actual tilt angles of less than 5° showed the projection symmetry clearly, owing to their unusual thickness of ~240 Å. After three cycles of unbending and correction for the contrast transfer function, amplitude and phase data of 60 images of tilt ranges 0° to 60° (Table 1) were merged. Two rounds of refinement of the tilt angle and tilt axis angle of the individual images against the merged data set resulted in an improvement of the average phase residual from 35 to 26°. Lattice lines (Fig. 1) were plotted with merged amplitudes and phases and manually truncated at points where the scatter of the observed phases approached random. The point-spread function indicated a nominal maximum resolution of 8 Å in the x-y plane and 22 Å in the perpendicular direction. A three-dimensional density map was calculated with the CCP4 program suite²⁵ with or without applying a negative temperature factor. The map was displayed with the graphics programs O (ref. 26) or AVS²⁷. Adjacent monomers related by non-crystallographic 6-fold symmetry were masked and averaged using the program SPIDER²⁸.

Received 23 December 1997; accepted 18 March 1998.

1. Scarborough, G. A. in *Handbook of Biological Physics* Vol. 2 (eds Konings, W. N., Kaback, H. R. & Lolkema, J. S.) 75–92 (Elsevier Science, Amsterdam, 1996).
2. Scarborough, G. A. in *Molecular Aspects of Transport Proteins* (ed. de Pont, J. J. H. M.) 117–134 (Elsevier, Amsterdam, 1992).
3. Nakamoto, R. K. & Slayman, C. W. Molecular properties of the fungal plasma-membrane H⁺-ATPase. *J. Bioenerg. Biomemb.* **21**, 621–632 (1989).
4. Nakamoto, R. K., Rao, R. & Slayman, C. W. Transmembrane segments of the P-type cation-transporting ATPases. *Ann. N. Y. Acad. Sci.* **574**, 165–179 (1989).
5. Mandala, S. M. & Slayman, C. W. The amino and carboxyl termini of the *Neurospora* plasma membrane H⁺-ATPase are cytoplasmically located. *J. Biol. Chem.* **264**, 16276–16281 (1989).
6. Hennessey, J. P. Jr & Scarborough, G. A. Direct evidence for the cytoplasmic location of the NH₂- and COOH-terminal ends of the *Neurospora crassa* plasma membrane H⁺-ATPase. *J. Biol. Chem.* **265**, 532–537 (1990).
7. Scarborough, G. A. & Hennessey, J. P. Jr Identification of the major cytoplasmic regions of the *Neurospora crassa* plasma membrane H⁺-ATPase using protein chemical techniques. *J. Biol. Chem.* **265**, 16145–16149 (1990).
8. Rao, U. S., Hennessey, J. P. Jr & Scarborough, G. A. Identification of the membrane-embedded regions of the *Neurospora crassa* plasma membrane H⁺-ATPase. *J. Biol. Chem.* **266**, 14740–14746 (1991).
9. Rao, U. S., Bauzon, D. D. & Scarborough, G. A. Cytoplasmic location of amino acids 359–440 of the *Neurospora crassa* plasma membrane H⁺-ATPase. *Biochim. Biophys. Acta* **1108**, 153–158 (1992).
10. Goormaghtigh, E., Chadwick, C. & Scarborough, G. A. Monomers of the *Neurospora* plasma membrane H⁺-ATPase catalyse efficient proton translocation. *J. Biol. Chem.* **261**, 7466–7471 (1986).
11. Chadwick, C., Goormaghtigh, E. & Scarborough, G. A. A hexameric form of the *Neurospora crassa* plasma membrane H⁺-ATPase. *Arch. Biochem. Biophys.* **252**, 348–356 (1987).
12. Henderson, R. & Unwin, P. N. T. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257**, 28–32 (1975).
13. Henderson, R. et al. Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213**, 899–929 (1990).
14. Kühlbrandt, W., Wang, D. N. & Fujiyoshi, Y. Atomic model of plant light-harvesting complex by electron crystallography. *Nature* **367**, 614–621 (1994).
15. Walz, T. et al. The three-dimensional structure of aquaporin-1. *Nature* **387**, 624–627 (1997).
16. Cheng, A., v. Hoek, A. N., Yeager, M., Verkman, A. S. & Mitra, A. Three-dimensional organisation of a human water channel. *Nature* **387**, 627–630 (1997).
17. Li, H., Lee, S. & Jap, B. Molecular design of aquaporin-1 water channel as revealed by electron crystallography. *Nature Struct. Biol.* **4**, 263–265 (1997).

18. Cyrklaff, M., Auer, M., Kühlbrandt, W. & Scarborough, G. A. 2-D structure of the *Neurospora crassa* plasma membrane ATPase as determined by electron cryomicroscopy. *EMBO J.* **14**, 1854–1857 (1995).
19. Goormaghtigh, E., Vigneron, L., Scarborough, G. A. & Ruysschaert, J.-M. Tertiary conformational changes of the *Neurospora crassa* plasma membrane H⁺-ATPase monitored by hydrogen/deuterium exchange kinetics: a Fourier transform infrared spectroscopy approach. *J. Biol. Chem.* **269**, 27409–27413 (1994).
20. Toyoshima, C., Sasabe, H. & Stokes, D. L. Three-dimensional cryo-electron microscopy of the calcium ion pump in the sarcoplasmic reticulum membrane. *Nature* **362**, 469–471 (1993).
21. Serrano, R. Structure and function of proton translocating ATPase in plasma membranes of plants and fungi. *Biochem. Biophys. Acta* **947**, 1–28 (1988).
22. Scarborough, G. A. Large single crystals of the *Neurospora crassa* plasma membrane H⁺-ATPase: An approach to the crystallisation of integral membrane proteins. *Acta Crystallogr. D* **50**, 643–649 (1994).
23. Dubochet, J. et al. Cryo-electron microscopy of vitrified specimens. *Quart. Rev. Biophys.* **21**, 129–228 (1988).
24. Crowther, R. A., Henderson, R. & Smith, J. M. MRC image processing programs. *J. Struct. Biol.* **116**, 9–16 (1996).
25. Collaborative Computational Project No. 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
26. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps. *Acta Crystallogr. A* **47**, 110–119 (1991).
27. Sheehan, B., Fuller, S. D., Piquis, M. E. & Yeager, M. AVS software for visualisation in molecular biology. *J. Struct. Biol.* **116**, 99–106 (1996).
28. Frank, J. et al. SPIDER and WEB: Processing and visualisation of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
29. Kühlbrandt, W. Discrimination of protein and nucleic acids by electron microscopy using contrast variation. *Ultramicroscopy* **7**, 221–232 (1982).
30. Unwin, P. N. T. & Ennis, P. D. Two configurations of a channel-forming membrane protein. *Nature* **307**, 609–613 (1984).

Acknowledgements. We thank M. Radermacher for help with the program SPIDER, D. Mills and J. Vonck for help with the JEOL 3000 SFF, and J. Postma for assistance with computer graphics. W.K. gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft. G.A.S. was supported by a grant from USPHS NIH.

Correspondence and requests for materials should be addressed to W. K. (e-mail: kuehlbrandt@biophys.mpg.de).

corrections

The effect of sedimentary cover on the flexural strength of continental lithosphere

Luc L. Lavier & Michael S. Steckler

Nature **389**, 476–479 (1997)

The last term on the right-hand side of equation (1) should read

$$\left(\frac{\epsilon}{A}\right)^{1/n} \int_{z_2}^{z_3} e^{u(z)} dz$$

to make it consistent with an integration over the variable z rather than over the function $u(z)$. □

Requirement for IRF-1 in the microenvironment supporting development of natural killer cells

Kouetsu Ogasawara, Shigeaki Hida, Nazli Azimi, Yutaka Tagaya, Takeo Sato, Taeko Yokochi-Fukuda, Thomas A. Waldmann, Tadatsugu Taniguchi & Shinsuke Taki

Nature **391**, 700–703 (1998)

In the second sentence of the second paragraph on page 701, interleukin-5 should be interleukin-15. □

A column on chromatography

This edition offers information on affinity chromatography media, a wedge-shaped column, instrument control software, and a variety of hyphenated systems for separation and analysis.

Electropak columns

From HPLC Technology

Capillary electrochromatography (CEC) columns created using electrokinetic packing

Produced by Unimicro Technologies, these columns are all individually tested under CEC conditions with measured peak efficiencies of up to 400,000 plates per metre and column-to-column reproducibility of retention of <2 per cent RSD for test solutes. Columns are packed with high-quality 3- μ m spherical silicas and are available in a range of functionalities, including C18, C8, phenyl, nitrile and ion exchange. (A 2- μ m C 18 resin is also available.) For CEC operation there are 75- μ m and 50- μ m internal diameter fused-silica columns, which are available in a range of packed column lengths from 20–40 cm, making full use of the high measured efficiencies found using this separation technique. Longer column packed lengths are possible as no pressure is produced either during the packing process or during the operation of these columns.

Reader Enquiry No. 100



Right in line, Supelco's Discovery C18 columns.

ducibility. A six-page certificate of analysis is included with each column, describing more than 50 specifications and tests passed by the packing material and/or column. A free 30-day evaluation of the HPLC column is offered by the company.

Reader Enquiry No. 102

Wedge column

From Sepragen

The wedge shape is designed to tackle problems found in scale-up operations

These columns take advantage of the separation capability and high capacity of low-cost soft media, as well as the characteristics of faster flow media, without compromising performance in terms of chromatographic separation efficiency and productivity. The Wedge column combines the design of 'radial flow' and traditional axial flow column chromatography. This permits easy linear scale-up of processes from laboratory to production scale. The chromatographic process can be transferred from a 10-ml axial column and the process will be optimized. The process can then be directly applied to a large-scale radial flow column: 10 l, 50 l, 100 l or even 300 l. This is possible as the column is just a section out of a process radial column.

Reader Enquiry No. 103

Affinity chromatography media

From Sigma

A selection of over 500 affinity resins suitable for almost any biopurification process

This product line includes many combinations of matrices, ligands and methods of attachment. For ease of selection, the affinity resins are organized by functional groups. The product categories include activated matrices/functionalized, amino acid, avidin/biotin, carbohydrate, dye, glutathione, hydrophobic, immunochemical, lectin, nucleic acid resins, and nucleotide/coenzyme resins, as well as specialty resins. The specialty group includes protein A and protein G matrices, and resins with applications in protein analysis, characterization and capture.

Reader Enquiry No. 101

Discovery C18

From Supelco

A C18 HPLC column designed, manufactured and tested for drug discovery and pharmaceutical method development

Based on a silica with carefully controlled pore characteristics, ultra-low metal content and high mechanical and chemical stability, these columns have been designed to ensure good peak shape for basic drugs and metal chelating compounds, long column life, and excellent column-to-column repro-

method to resolve the unwanted enantiomers of amino acids by both capillary GC and LC — a process that does not require derivatization. Many of the methods for enantiomer separation deal strictly with primary α -amino acids and cannot accommodate either β -amino acids or secondary amines like proline. The macrocyclic glycopeptide, teicoplanin, is bonded to a high-purity silica gel to produce a chiral stationary phase. Designated Chirobiotic T, it is said to be a highly effective analytical and preparative media for the resolution of underivatized amino acid enantiomers. According to the manufacturer, it can separate all 20 naturally occurring amino acids enantiomers and over 100 other α and β amino acids. An advantage to using this technique is the ability to resolve these amino acids in simple alcohol/water mixtures, allowing low ultraviolet detection. For the multifunctional amino acids like aspartic/glutamic or lysine/arginine, it is only necessary to add enough acetic acid to the mobile phase to reach a pH of 3.8. Many N-blocked amino acids can be separated as well, some requiring the use of the complementary stationary phase, Chirobiotic V.

Reader Enquiry No. 104

Tools and systems

Chrom Link C/S

From Beckman

An analog interface connecting chromatographic instruments to the PeakPro C/S chromatography data system

The analog module interfaces to any analog output chromatograph, with up to two independent instruments per unit. It has a minimum 25-bit resolution at sampling rates of 120 Hz, increasing to 31 bits at 2 Hz. The instrument's interface combines an alphanumeric keyboard and graphic liquid crystal display to make it a complete data station in the laboratory. The Chrom Link C/S uses a high-speed microprocessor for fast data reduction and buffering before



Missing link — Beckman's analog interface.

Chirobiotic T

From Astec

An HPLC method for the separation of underivatized amino acids

This is a new, simplified approach that should prove effective for the separation of a broad range of amino acids and small peptides, using the macrocyclic, glycopeptide-bonded stationary phase. Astec has developed this

sending the data to the host computer through a direct Ethernet link. The unit uses four megabytes of dynamic RAM memory for data buffering, which should provide hours of uninterrupted operation in case the host system goes down; and, when communication with the host is restored, it sends the data automatically. The Chrom Link C/S also provides inputs and outputs for interfacing to autosamplers and switching external valves.

Reader Enquiry No. 105

Xcalibur

From Finnigan

A flexible data system that presents a unified interface and feature set for a range of GC/MS and LC/MS users

This is a second generation Windows NT platform that provides complete instrument control of Finnigan mass spectrometers and other ThermoQuest instruments, including the XSQ 7000 series, GCQ, LCQ, MD800, Voyager, Navigator and AutoMass III. It provides full control for the new Trace GC and SpectraSystem LC modules and the UV6000 photo-diode-array detector, as well as other third-party systems. The software is said to provide a universal system that allows chromatographers and mass spectrometrists to learn a single system for instrument set up, processing, on-line library searches, and more. Users can import data from old systems and add new instruments as laboratory requirements grow. The core components of this system include QuanBrowser, QualBrowser, NIST Library, Custom Reports, which are enhanced by a comprehensive set of optional plug-in application modules, such as BioWorks (featuring Sequest), LCQuan and Discovery for combinatorial chemistry.

Reader Enquiry No. 106

ÅKTA[®]PLC

From Amersham Pharmacia Biotech

Laboratory-scale protein purification (from µg to mg) with Windows NT software control

This benchtop separation system has been optimized for use with all aqueous chromatographic techniques, such as gel filtration, ion exchange, affinity and hydrophobic interaction chromatography. It is said to be fully compatible with the high salt concentrations typically used in ion exchange, as well as allowing flow rates of up to 20 ml min⁻¹ to be used routinely. The unit has a built-in gate that functions as a rack and holds a closed configuration for compactness. Accessories are held in place by a simple clip, facilitating easy addition and removal. An extensive set of fully tested method templates is provided with the system, and an 'adviser' is included to give instant access to advice and technical information on all aspects of chromato-



Pronunciation may be the most difficult part of using the ÅKTA[®]PLC chromatography system.

graphy. A comprehensive column library is also provided, which automatically sets run parameters into the method template, enabling simple, safe column selection and enhancing run reproducibility. Both ultraviolet absorption and conductivity flow cells are supplied as standard, with a pH flow cell as an optional extra. For ultraviolet detection, 254–280-nm filters are supplied as standard and additional options enable the selection of a wider range of wavelengths. All of the components in the system are biocompatible. Moreover, the latest version of Unicorn software, for WindowsNT, has been developed to aid in compliance with good laboratory and manufacturing practices, and offers better networking compatibility and improved multi-tasking.

Reader Enquiry No. 107

ABI 165 LC/MS

From PE Sciex

A rugged and reliable high-performance, single-quadrupole liquid chromatography, mass spectrometry (LC/MS) system

This system is said to deliver accurate molecular weights at high sensitivity and resolution. It features a small footprint and a new atmospheric pressure ionization (API) interface with PE Sciex's curtain gas for ruggedness, reliability, ease of use and optimization. The system's mass range handles analyses from small molecules to large proteins, from routine combinatorial chemistry and pharmacology to protein and peptide characterization. There is a choice of inlets ranging from ion spray to APCI (heated nebulizer), and the system is capable of analysing a wide range of compounds with varying structures, polarities and molecular weights, says the manufacturer. The system uses MSeXpress software, as well as the new MassChrom data handling software, which incorporates a suite of integrated programs for data acquisition, manipulation, quantification and reporting. This software also offers flexible automation through AppleScripts to fully automate both run-time and post-run processing tasks, ranging from the simple to the complex.

Reader Enquiry No. 108

175 GC-UV

From Inscan

A new instrument that combines gas chromatography with an ultraviolet spectrophotometer

This instrument scans molecular separations in the 168–330-nm range and is enhanced by software and an extensive reference library. It enables a wide variety of compounds to be identified and quantified quickly, especially in the important spectral region between 168 and 190 nm. Inscan 175 GC-UV is designed to be a versatile instrument that can analyse solid, liquid and gas samples in the same apparatus. According to the manufacturer, it is easy to operate, requires a minimum of bench space and has no moving parts. The instrument has found applications in a variety of applications, including the identification of isomers, analyses of cigarette smoke, dust particles and aircraft fuel, to name a few. Air pollutants in the pulp and paper industry and breath from diabetic patients have also been studied. It may prove useful as an all-round instrument in analytical and environmental laboratories engaged in raw material analysis, process development studies or basic molecular research.

Reader Enquiry No. 109

Connex

From J&W Scientific

A new column connector technology that is designed to reduce column installation time

According to the manufacturer, Connex is an improvement over metal, fused silica and glass connectors. These capillary column connectors are said to provide a leak-free, inert and reusable junction that can be connected or disconnected instantly. There is said to be no time-consuming measuring, or messy gluing. Connex kits have been designed initially for use with Hewlett-Packard GC and GC/MS systems. Kits for Varian, Perkin-Elmer and Shimadzu components will soon follow.

Reader Enquiry No. 110

Instrument control and data systems

Millennium32

From Waters

A 32-bit software application for gas and liquid chromatography data analysis

This package is the third release in the series. However, it is not a revision of the earlier 16-bit software version 2.1, but rather, an entirely new 32-bit software product. Designed expressly for Windows 95 and Windows NT 4.0, it takes advantage of and replicates the popular look, feel, functionality and ease of use of these operating environments and their user interfaces. In addition, the software's

architecture cost-effectively transforms the ways in which laboratories respond to regulatory and legal requirements in product manufacture, says Waters. The software provides the protected-mode, security of application integrity and operation of a Windows environment. It is said to be easier to learn and use than earlier versions, without sacrificing any of the efficacy of its predecessors. In fact, several new capabilities make Millennium32 software well suited for the total analysis of data in both liquid and gas chromatography. These include tutorial Wizards and extensive context-sensitive online help functions, enhanced security with advanced privilege processing, view filters that permit a high degree of customization to display and manage data, and improved capabilities in generating, configuring and securing audit trails.

Reader Enquiry No. 111

Star 5.0

From Varian

Workstation software that runs in the Windows95 or Windows NT 4.0 environments

Harnessing the 32-bit computing environment, this system should help with single, manual sample injection or an automated sequence of 100 samples. According to Varian, the workstation can be up and running at full speed within minutes and uses 'configuration wizards' to walk the user through the setup of hardware, method building steps that take the user through new chromatographic method construction and simplified automation controls to make it easy to run large batches of samples. The system includes password protection to prevent accidental alteration of important methods and data, and security administration that disallows unwanted interference on instruments during unattended operation.

Reader Enquiry No. 112

Atlas

From Thermo LabSystems

A chromatography data system for use in large and small laboratories

This new suite uses 32-bit data handling for chromatographic analysis, mirroring the chromatographers own workflow in laboratory environments. The company also features its LIMS solutions, such as the 32-bit SampleManager 4.0 for Windows NT. This enterprise-level laboratory information management system is available to satisfy the needs of large multi-user customers in demanding process and pharmaceutical applications. New industry-specific LIMS modules include Stability 3.2, which is designed for shelf-life analysis in regulated and vali-



Thermo LabSystems' Atlas uses Windows.

dated environments, such as the pharmaceutical industry, SQC 3.2 for quality control and SM-IDI for the approved SAP R/3 interface.

Reader Enquiry No. 113

Chromeleon

From GynkoteK

A data system that offers secure network operation and facilitates documentation of all instrument parameters

The data system provides all working steps in method development and data reprocessing. A tutorial is now available in three languages to enhance learning and familiarity with this package. By importing worklists of a superordinate LIMS, Chromeleon can be smoothly integrated in existing software environments. A new user management module handles data and system access within network environments. 'Access groups' allow user access to be defined to systems and datasources, whereas 'privilege groups' serve to create user profiles with 70 individually assignable privileges, defining the modifications the user groups can perform. To further enhance data security, 'locking' data directories is now possible, thus protecting the data from any changes. Comprehensive documentation requirements are fulfilled by the newly integrated 'change history' function. The system records all changes to user data. If the user moves a peak delimiter during re-integration, detailed information on this action (user identification, date/time, peak delimiter [min] before/after) is automatically recorded in the sample history.

Reader Enquiry No. 114

HP ChemStation

from Hewlett Packard

Instrument control for GC, LC, LC/MS, CE, ultraviolet-visible spectroscopy and analog-to-digital conversion across LANs

This version of HP ChemStation includes several quality improvements and performance enhancements to help chemical and pharmaceutical analysts increase productivity. The most significant of the many new features is the ability to control instruments

across local area networks (LANs) and the ability to publish analysis reports on intranets. Using the latest LAN technologies, the HP ChemStation allows analysts to control instruments from any PC and to break the physical distance limitations of conventional communication interfaces. Now the PC no longer needs to be next to the instrument, saving bench space in the laboratory and allowing analysts to work from the comfort of their offices. The revision of ChemStation also includes an HTML export feature, which enables analysts to create chromatographic reports in Web-page format ready for publishing on their intranets.

Reader Enquiry No. 115

Literature

1998 catalogue series

From Fisher Scientific

Three comprehensive catalogues covering products for the general science laboratory, chemicals and chromatography

The series covers over 2,400 pages and nearly 30,000 products. The general scientific catalogue of apparatus, equipment and consumables features a significant range of new Fisher products, new tissue and cell culture products from Nunc and Corning Costar, Eppendorf liquid handling and Lancer washing machines. The chemicals catalogue presents over 8,000 products, including new and improved specifications on HPLC solvents, new scintillation cocktails from Wallace and new pesticide standards from LGC. The catalogue also includes the availability of enhanced technical information in several specialist areas. The chromatography catalogue includes Mallinckrodt Baker SPE products, providing a variety of choices to the end user.

Reader Enquiry No. 116

Model 7955 column heater

From Jones Chromatography

Literature covering the 7955 HPLC column heater/chiller

The manufacturer offers a new four-colour, two-page brochure that outlines the benefits of this HPLC column heater/chiller. This is a solid-state heat exchanger that controls temperatures at ambient, making it useful for GLP and FDA regulated analyses, says the company. The cooling unit does not require external services, such as water or compressed air, and displays temperature on a large liquid crystal display. The brochure includes additional unit specifications and ordering information.

Reader Enquiry No. 117

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.